Traffic equilibrium between transit lines serviced by capacitated vehicles: a route choice model with passenger waiting on platform

Fabien Leurent

Traffic equilibrium between transit lines serviced by capacitated vehicles: a route choice model with passenger waiting on platform

Fabien Leurent ([1])
Université Paris Est, Laboratoire Ville Mobilité Transport, Ecole des Ponts ParisTech

Abstract

A theory is provided for the problem of passenger waiting and route choice at a transit station where a set of lines serviced by vehicles of limited capacity are available to reach a given destination. It is assumed that the platform layout involves one boarding zone only. The theory addresses waiting discipline – either priority queuing or mingled waiting – and choice behaviour – either selfish or social.

The problem of passenger traffic assignment to capacitated transit lines is modelled in a stationary framework, basically assuming memoryless services and passenger arrivals. The size of the passenger stock on the platform is the main state variable as it determines the attractivity condition of a line with respect to alternative routes. Traffic equilibrium is based on the attractivity of a line bundle; it is characterized as the solution to a recursive program. Its existence and uniqueness are demonstrated. When the stock size increases the attractive bundle is enlarged and its average cost is increased. An efficient solution algorithm is provided. The transition from traffic theory to network assignment is discussed. Lastly, a Markovian model of the traffic problem is developed; an analytical solution is given for the case of two lines of unit or infinite vehicle capacity.

Keywords

Transit vehicle capacity. Station platform. Passenger stock. Attractivity threshold. Priority queuing. Mingled waiting. Selfish behaviour. Social behaviour

1. Introduction

Background. A transit mode of transportation involves the boarding of passengers in service vehicles at station nodes, prior to carrying them aboard up to their alighting station. At a station platform, the passenger has to wait for a vehicle to arrive. When the stock of waiting passengers exceeds the capacity available in the vehicle, then some passengers have to wait further. Reducing the users' waiting time is a stake of paramount importance to the network operator who strives to deliver a satisfactory quality of service. Two families of models are available to design service plans: first, 'bulk' models in queuing theory (e.g. Kleinrock, 1975) have some analytical properties of limited practical value; second, in traffic assignment theory a series of models have been developed to address passenger route choice onto a transit network, eventually considering capacity constraints. The dominant model for uncapacitated route choice at a station to a given destination is due to Chriqui and Robillard (1975): it has been generalized to network assignment by Spiess and Florian (1989), who also addressed the

[1] Author's address: fabien.leurent@enpc.fr. Ecole des Ponts ParisTech, Laboratoire Ville Mobilité Transport, 6-8 avenue Blaise Pascal, Champs sur Marne 77455 Marne la Vallée Cedex

issue of on-board crowding discomfort. De Cea and Fernandez (1993) linked the line frequency at the station level to the flows of passengers either on egress, access or board. Cominetti and Correa (2001) further specified the frequency function with respect to passenger flow by traffic stream. Cepeda et al (2006) generalized these principles in a network assignment model. Other models have been proposed: a recent review is available in Leurent and Askoura (2010).

A major difference between queuing models and assignment models is that the former involve the number of passengers waiting on platform as the main state variable, whereas the latter ignore it – at least in their static version. Indeed, stock variables have been neglected in static traffic assignment so far, despite their recognition in queuing theory which is, basically, stationary in nature.

Objective. This paper addresses the effect of limited vehicle capacity onto passenger waiting and route choice at a transit station. It brings about a model of passenger traffic assignment to transit lines from a station platform to a given destination. Two scales of analysis are integrated, microscopic and macroscopic.

The microscopic scale pertains to the route choice behaviour of the individual network user in relation to the route options, of which the features make up the quality of service that determines route choice: here the focus is on the conditions of passenger waiting for service vehicles in terms of individual waiting time and the priority rank to avail oneself of some place within a vehicle at dwelling.

At the macroscopic scale, the focus is on the quantitative relationship between vehicle capacity and frequency of operation by transit line, on one hand, and the passenger flow arriving at the station to get to a given destination, on the other. These macroscopic factors, combined with the passenger economic behaviour of route choice and the waiting discipline, determine the line attractivity and the size of the passenger stock waiting on the origin platform. Thus the interplay of microscopic and macroscopic scales induces a state of traffic equilibrium between passenger demand and line services.

The scope of the model is limited to a single origin-destination pair and passengers that are homogeneous save for their instant of arrival in the system. Despite this limitation, the model indicates the effect of vehicle capacity on passenger waiting at the platform and the split of passenger flow between the available transit lines, yielding insight into physical phenomena that involve microeconomic behaviour.

Approach. The analysis is conditional on the stock size and proceeds along the following track. First, there is the issue of vehicle capacity: when a vehicle arrives at the station, would its capacity suffice to empty the stock? This is the primary concern if there is one line only. Line attractivity is the second and major issue: if several lines link the transit station to the destination, then some of them may be less attractive and their vehicles might not be used to their full capacity even if it is less than the stock size, should some better ranked passengers prefer to wait in order to avail themselves later of a faster line. To each line is associated an *attractivity threshold* that is the maximum stock up to which the line is unattractive: the number of passengers boarding a vehicle is the minimum of not only vehicle capacity but also the rest of stock size minus attractivity threshold.

The third issue pertains to station layout and waiting protocol. It is assumed that there is one platform with one waiting zone only. Two waiting protocols are considered, either priority ranking in a First In First Out manner (save for the access to less attractive lines) versus mingled waiting (without priority). Under priority queuing the analysis is based on passenger rank which indicates the size of that part of the stock that holds more priority than the passenger

Fourth and last is addressed the issue of route choice rationality, either selfish individual behaviour that leads to user equilibrium or social behaviour that leads to system optimization.

Contribution. By jointly considering vehicle capacity, line attractivity, waiting discipline and type of rationality, based on passenger flow rate and line frequency, conditions for traffic equilibrium of passenger demand and line services are stated formally as an attractive bundle of lines, i.e. an optimal travel strategy in the words of Spiess and Florian (1989). The attractive bundle depends on stock size (or priority rank). The attractivity conditions are characterized by a mathematical program, yielding properties of existence and uniqueness of traffic equilibrium. Any line will belong to the attractive bundle from its attractivity threshold. An efficient algorithm is provided, which extends that of Chriqui and Robillard (1975) to capacitated assignment. On assuming large values for the size variables, line choice probabilities are derived that involve line capacity i.e. service frequency times vehicle capacity, instead of line frequency only. There remains the issue of which size variable would correspond to a given origin-destination flow. Our model yields the inverse relationship as it derives the exit flow from the stock size. Application to network assignment is addressed in a subsequent paper (Leurent, 2009b). Here a Markovian queuing model is developed that embodies the full set of assumptions: analytical solutions are provided for simple binary cases, with an explicit relationship between OD flow and average size.

Paper outline. The rest of the paper is structured in six sections. Section 2 recalls the uncapacitated model and states the basic framework of a capacitated model: a binary instance is used to address the issues of waiting protocol and rationality type. Section 3 develops the capacitated model under priority queuing on the basis of an induction principle that pertains to the size variable. The concepts of relative capacity, composed costs, attractivity threshold and attractive bundle are introduced. Then traffic equilibrium is stated formally and characterized by a mathematical program. Structural properties are stated and the assignment algorithm is provided. Section 4 develops a parallel analysis for the capacitated model under mingled waiting. Section 5 discusses the results in the perspectives of traffic theory and assignment models. Section 6 brings about the Markovian queuing model of line attractivity, with explicit analytical solution of simple binary cases. Lastly, Section 7 concludes and points to potential developments.

2. The capacitated attractivity problem

Let us consider a transit station at which a set Z of lines z are available to reach the destination. By assumption, line z is serviced at operation frequency f_z by vehicles of homogeneous passenger capacity k_z. Denote by λ the flow rate of passenger arrivals at the origin station. If only one line is available then queuing is a simple phenomenon that can be modelled as a traffic bottleneck in a macroscopic perspective. However, if several lines are available there arises the issue of line attractivity: should a passenger prefer to use a line of which one vehicle is dwelling and immediately available or to wait for another line presumably faster?

This section is purported to explore the relationship between capacity constraints and line attractivity. After recalling the model of attractivity in an uncapacitated system (§2.1), some generic properties are stated for a capacitated system about the assignment of passengers to attractive lines (§ 2.2). Then, specific properties about the individual waiting time and travel time are established under, in turn, priority queuing (§ 2.3) and mingled waiting (§ 2.4):

throughout the investigation of specific properties, the statement of principles is illustrated using a binary instance. Next, the transition from selfish behaviour to system optimization is investigated (§ 2.5). Lastly, the influence of platform layout and management by the station operator is discussed (§ 2.6).

TABLE OF NOTATION

Z (resp. A)	Set (resp. subset, bundle) of available lines denoted as z, a
t_a	Run time of line a from platform to destination
f_a	Operation frequency of line a
k_a	Supplied passenger capacity in a vehicle of line a
α	Specific discomfort factor of wait time relative to run time
w_a	Wait time at platform for a vehicle a to arrive
g_a (resp. g_A)	Generalized travel time of line a (resp. of line bundle A)
n	Size of passenger stock waiting on platform
$\bar{t}_a^{(n)}$	Composed time of line a
$k_{a/n}$	Attractive capacity of a line a vehicle, wiith respect to stock size n
N_a	Attractivity threshold of line a
$\overline{\theta}_n$	Cost (generalized travel time) of reference strategy
$\overline{\theta}_n^{-a}$	Generalized time of strategy alternative to a
$A(n)$	Attractive bundle at order n
$\mathrm{N}^*(\zeta)$	Generating function of stock size
q_{nm}	Transition rate from state n to state m
λ	Flow rate of passenger arrivals at station
φ	$= f_a / \lambda$
ρ	Parameter of a geometric sequence
π_n	Stationary probability of state n
$\bar{x}_a$	Average passenger boarding in line a

2.1 On attractivity in an uncapacitated system

When a vehicle of line z is dwelling at the station, every potential user may evaluate the opportunity of using it by comparing its planned time to destination, say t_z, to the time that another line a can be expected to deliver. Assumedly the alternative line is not available at the instant of choice: if its operation is memoryless then the average waiting time is $1/f_a$, yielding an expected travel time of $\bar{g}_a = t_a + f_a^{-1}$ or more generally $\bar{g}_a = t_a + \alpha f_a^{-1}$ if the user associates a discomfort cost of α to a unit wait time as compared to a unit in-vehicle time.

Thus the attractivity of service z requires that

$$t_z \le \bar{g}_a,\ \forall a \in Z - \{z\}. \tag{2.1}$$

The necessary condition (2.1) does not suffice to characterize the set of attractive lines, say A, if more than two lines are eligible, in which case each line z at an instant of availability is faced to a set of alternatives so the waiting time for any of them is the minimum of their respective waiting times. Denoting by $w_B \equiv \min\{w_z : z \in B\}$, the alternative travel time is:

$$g_{A-z} = \alpha w_{A-z} + \sum_{a \in A-z} 1_{\{w_a = w_{A-z}\}} t_a. \tag{2.2}$$

On the average,

$$\bar{g}_{A-z} \equiv \mathrm{E}[g_{A-z}] = \alpha \bar{w}_{A-z} + \sum_{a \in A-z} \Pr\{w_a = w_{A-z}\} t_a. \tag{2.3}$$

Thus the necessary and sufficient condition for attractivity is that

$$z \in A \ \text{ if and only if } \ t_z \le \bar{g}_{A-z}. \tag{2.4a}$$

Chriqui and Robillard (1975) showed that this is equivalent to $t_z \le \bar{g}_A$ when the services are delivered in a Markovian (memoryless) process. Let us demonstrate shortly this property.

Lemma 0, optimization of a line bundle. *Let $B \subset Z$ and $z \in B$ with $f_z > 0$: it holds that*

$$t_z \gtrless \bar{g}_B \ \Leftrightarrow\ \bar{g}_B \gtrless \bar{g}_{B-z}. \tag{2.4b}$$

Proof. Decompose $f_B \bar{g}_B^{(n)} = \alpha + f_z \bar{t}_z^{(n)} + \sum_{a \in B-z} f_a \bar{t}_a^{(n)}$ so $f_{B-z}\bar{g}_B = f_{B-z}\bar{g}_{B-z} + f_z(t_z - \bar{g}_B)$. Then $\bar{g}_B = \bar{g}_{B-z} + (t_z - \bar{g}_B) f_z / f_{B-z}$. By the positivity of the frequencies, $\bar{g}_B \ge \bar{g}_{B-z}$ iff $t_z \ge \bar{g}_B$, and $\bar{g}_B \le \bar{g}_{B-z}$ iff $t_z \le \bar{g}_B$.

In the absence of any capacity constraint, every user among the passenger stock of size say n can take a vehicle of an attractive line when it becomes available. Whatever n, the probability to take line a is proportional to its frequency of operation, yielding a modal share of

$$\eta_a^{(n)} = \frac{f_a}{f_A}, \text{ wherein } f_A \equiv \sum_{a \in A} f_a. \tag{2.5}$$

As this does not depend on n, neither does the average modal share of the line, so that $\bar{\eta}_a = \eta_a^{(n)}$.

The minimum wait time of a user for an attractive line is w_A. Under the assumption of independent, Markovian services, each random variable w_z is distributed exponential with parameter f_z, so the statistical independence implies that w_A is also an exponential random variable with parameter equal to the sum of its operands' ones i.e. f_A. Thus the average wait time is:

$$\overline{w}_A \equiv \mathrm{E}[w_A] = 1/f_A . \tag{2.6}$$

Combining (2.3), (2.5) and (2.6) yields the expected travel time by using the attractive set of lines:

$$\overline{g}_A \equiv \mathrm{E}[g_A] = \frac{\alpha}{f_A} + \sum_{a \in A} \frac{f_a}{f_A} t_a . \tag{2.7}$$

Travel time as a random variable g_A has the same distribution as the independent sum of one exponential variable, w_A, multiplied by α and a mix of deterministic variables t_a with given proportions f_a / f_A. Therefore, its variance is the sum of that of αw_A and the variance of the mix which is reduced to its interclass part:

$$\mathrm{V}[g_A] = \frac{\alpha^2}{f_A^2} + \sum_{a \in A} \frac{f_a}{f_A} (t_a - \bar{t}_A)^2 , \tag{2.8}$$

wherein $\bar{t}_A \equiv (\sum_{a \in A} f_a t_a) / f_A$.

2.2 Attractivity under capacity: generic properties

In reality, each vehicle for passenger transit has an on-board capacity that is limited. Assuming that the vehicles servicing line a have homogeneous passenger capacity of k_a, an integer number, then the capacity supplied during a unit time period amounts to $\kappa_a = f_a k_a$. When the station is serviced by a vehicle of an attractive line a, the number n_a of passengers that can board in it is limited by both vehicle capacity and the stock size, n:

$$n_a = \min\{k_a, n\} . \tag{2.9}$$

It would be naïve, however, to assume that all of the n waiting passengers would evaluate the attractivity in a homogeneous way, independently of their priority rank or the number $(n - k_a)^+$ of passengers that would eventually remain on platform. To integrate such dependencies, let us introduce the notion of *attractive capacity*, denoted as $k_{a/n}$, of a line a vehicle with respect to stock size n. Prior to further specification in Section 3, some basic properties must hold for attractive capacity:

$$k_{a/n} \leq \min\{n, k_a\} \text{ as a capacity,} \tag{2.10a}$$

$$k_{a/n} \geq 1 \text{ if } k_a \geq 1 \text{ and line is attractive at } n \geq 1 . \tag{2.10b}$$

Under state n, the modal share of line a becomes proportional to $k_{a/n}$:

$$\eta_a^{(n)} = \frac{f_a k_{a/n}}{\sum_{z \in A(n)} f_z k_{z/n}} \text{ if } a \in A(n) \text{ or } \eta_a^{(n)} = 0 \text{ otherwise.} \tag{2.11}$$

Notation $A(n)$ means that the attractive set depends on the current state of the system. The overall modal share of a line depends on the attractive capacity throughout the current states and the probability π_n of each state throughout time i.e. at any instant of potential arrival for a vehicle:

$$\overline{\eta}_a = \frac{1}{\overline{n}} \sum_{n \geq 1} \pi_n . n . \eta_a^{(n)} \text{ wherein } \overline{n} = \sum_{n \geq 1} \pi_n . n . \tag{2.12}$$

From these modal shares stems the average run time between station and destination:

$$\mathrm{E}[t_A] = \sum_{a \in A} \overline{\eta}_a t_a . \tag{2.13}$$

2.3 Specific properties under priority queuing

By priority queuing it is meant here that each passenger waiting on the station platform has a priority rank, m, which determines his behaviour of line choice that depends on the passenger with lesser ranks (meaning higher priority) but not on those with greater rank. Denote by N_a the maximum rank up to which line a is not attractive. When a vehicle of that line becomes available, it will be taken by those passengers ranking in position $m \in [N_a + 1,..N_a + k_{a/n}]$ if $n > N_a$ or by nobody if $n \leq N_a$.

Under that queuing discipline, the attractive capacity and the attractive set depend on the priority rank. The waiting time as a random variable is composed in a recursive way as follows:

$$w^{(m)} = w_{A(m)} + \sum_{a \in A(m)} 1_{\{w_a = w_{A(m)}\}} [1_{\{m - N_a \in [1, k_{a/m}]\}} . 0 + 1_{\{m > N_a + k_{a/m}\}} . w^{(m - k_{a/m})}] . \tag{2.14}$$

On the average,

$$\overline{w}^{(m)} = \overline{w}_{A(m)} + \sum_{a \in A(m)} \frac{f_a}{f_{A(m)}} 1_{\{m > N_a + k_{a/m}\}} . \overline{w}^{(m - k_{a/m})} , \tag{2.15}$$

under the convention that $\overline{w}^{(0)} = 0$. This formula clearly differs from the average wait time in the uncapacitated model, which is reduced to $\overline{w}_A$.

This yields the generalized time of travel as a composed random variable:

$$g^{(m)} = \alpha w_{A(m)} + \sum_{a \in A(m)} 1_{\{w_a = w_{A(m)}\}} [1_{\{m - N_a \in [1, k_{a/m}]\}} . t_a + 1_{\{m > N_a + k_{a/m}\}} . g^{(m - k_{a/m})}] . \tag{2.16}$$

On the average,

$$\overline{g}^{(m)} = \alpha \overline{w}_{A(m)} + \sum_{a \in A(m)} \frac{f_a}{f_{A(m)}} [1_{\{m - N_a \in [1, k_{a/m}]\}} . t_a + 1_{\{m > N_a + k_{a/m}\}} . \overline{g}^{(m - k_{a/m})}] . \tag{2.17}$$

Instance 1. Let us consider two lines a and b with run time of $t_a < t_b$, respectively, and a unit vehicle capacity. This could correspond to two companies of taxi (which indeed is a mode for public transportation). Assume that only line a is attractive for a small passenger stock $n \leq N_b$. The n-th customer must wait for n vehicles to arrive in order to get one, i.e. $g^{(n)} = t_a + \alpha \sum_{i=1}^{n} w_a^{(i)}$, yielding

$$\overline{g}^{(n)} = t_a + \alpha n f_a^{-1} . \tag{2.18}$$

Line b becomes attractive from $n = N_b + 1$ such that

$$\overline{g}^{(N_b)} = t_a + \alpha N_b f_a^{-1} \leq t_b < t_a + \alpha (N_b + 1) f_a^{-1} , \tag{2.19}$$

hence $N_b \le \frac{f_a}{\alpha}(t_b - t_a) < N_b + 1$, or equivalently

$$N_b = \mathrm{Int}[\frac{f_a}{\alpha}(t_b - t_a)]. \tag{2.20}$$

For $n > N_b$, $\overline{g}^{(n)} = f_Z^{-1}[\alpha + f_a t_a + f_b t_b] + \overline{g}^{(n-1)}$.

In particular, at $n = N_b + 1$, $\overline{g}^{(N_b+1)} = \alpha(1+N_b) f_Z^{-1} + \tau$, in which $\tau \equiv (f_a t_a + f_b t_b)/f_Z$.

For $n \ge N_b + 1$, $\overline{g}^{(n)} = \alpha f_Z^{-1} + \overline{g}^{(n-1)} = \alpha(n - N_b - 1) f_Z^{-1} + \overline{g}^{(N_b-1)} = \alpha n f_Z^{-1} + \tau$. (2.21)

For numerical illustration let us take $\alpha = 60$ min/h, $f_a = f_b =$ 10/h, $t_a =$ 20' and $t_b =$ 40'. Then $N_b = 3$ and $\tau = 30'$, yielding the following sequence of average travel times $[\overline{g}^{(n)} : n \ge 1]$ = {26', 32', 38', 39', 42', 45'… regular step +3'}.

It turns out that line b is attractive from $n = 4$, despite $\overline{g}^{(4)} = 39' < 40' = t_b$. This apparent paradox makes an obvious difference to the uncapacitated model.

2.4 Specific properties under mingled waiting

'Mingled waiting' is a better wording than 'Mingled queuing' to define a system such that every waiting passenger has an equal chance of boarding in a dwelling vehicle, with success probability of $k_{a/n}/n$ which is null if line a is unattractive at stock size n.

'Naïve mingled waiting' means that a waiting passenger evaluates the attractivity of an immediately available service, say a, with respect to other services not yet available by considering that the stock size remains at $n - k_{a/n}$ after the departure of the present vehicle, thus neglecting the eventual arrival of subsequent passengers on the platform. The random variable associated with this (myopically) estimated wait time satisfies that

$$\begin{aligned} w^{(n)} &= w_{A(n)} + \sum_{a \in A(n)} 1_{\{w_a = w_{A(n)}\}} [\frac{1}{n} \sum_{i=1}^{n} 1_{\{i \le k_{a/n}\}}.0 + 1_{\{i > k_{a/n}\}} w^{(n-k_{a/n})}] \\ &= w_{A(n)} + \sum_{a \in A(n)} 1_{\{w_a = w_{A(n)}\}} (1 - \frac{k_{a/n}}{n}) w^{(n-k_{a/n})} \end{aligned} \tag{2.22}$$

This yields an average wait time of

$$\overline{w}^{(n)} = \overline{w}_{A(n)} + \sum_{a \in A(n)} \Pr\{w_a = w_{A(n)}\} \frac{n - k_{a/n}}{n} \overline{w}^{(n-k_{a/n})}_{A(n-k_{a/n})}. \tag{2.23}$$

The estimated travel time satisfies that

$$\begin{aligned} g^{(n)} &= \alpha w_{A(n)} + \sum_{a \in A(n)} 1_{\{w_a = w_{A(n)}\}} [\frac{1}{n} \sum_{i=1}^{n} 1_{\{i \le k_{a/n}\}} t_a + 1_{\{i > k_{a/n}\}} g^{(n-k_{a/n})}] \\ &= \alpha w_{A(n)} + \sum_{a \in A(n)} 1_{\{w_a = w_{A(n)}\}} \frac{k_{a/n} t_a + (n - k_{a/n}) g^{(n-k_{a/n})}}{n} \end{aligned}. \tag{2.24}$$

Thus the average travel time is

$$\overline{g}^{(n)} = \alpha \overline{w}_{A(n)} + \sum_{a \in A(n)} \frac{f_a}{f_{A(n)}} \frac{k_{a/n} t_a + (n - k_{a/n}) \overline{g}^{(n-k_{a/n})}}{n}. \tag{2.25}$$

Instance 1 (continued). Let us now assume that waiting passengers are mingled. As long as only line a is attractive, the average travel time satisfies that

$$\overline{g}^{(n)} = \alpha\overline{w}_a + [\tfrac{1}{n}t_a + \tfrac{n-1}{n}\overline{g}^{(n-1)}], \tag{2.26}$$

Hence
$$n\overline{g}^{(n)} - (n-1)\overline{g}^{(n-1)} = t_a + n\alpha / f_a . \tag{2.27}$$

Summing that from 1 to n and dividing by n yields

$$\overline{g}^{(n)} = t_a + \frac{\alpha}{f_a}\frac{n+1}{2}. \tag{2.28}$$

This result is intuitive as it would take n vehicles to service n passengers, yielding average total wait time of $\overline{w}_a \sum_{i=1}^{n} i = \overline{w}_a n(n+1)/2$ to be shared equally between the waiting passengers (in their myopic estimation).

Line b becomes attractive from $n = N_b + 1$ such that

$$\overline{g}^{(N_b)} = t_a + \tfrac{1}{2}\alpha N_b f_a^{-1} \le t_b < t_a + \tfrac{1}{2}\alpha(N_b+1)f_a^{-1}, \tag{2.29}$$

yielding

$$N'_b = \mathrm{Int}[2\tfrac{f_a}{\alpha}(t_b - t_a)]. \tag{2.30}$$

At $n > N'_b$, $\overline{g}^{(n)} = \alpha f_Z^{-1} + \frac{1}{n}\tau + (1-\frac{1}{n})\overline{g}^{(n-1)}$, or, equivalently,

$$n\overline{g}^{(n)} - (n-1)\overline{g}^{(n-1)} = \tau + n\alpha f_Z^{-1}. \tag{2.31}$$

Summing that over n from $N'_b + 1$ to n yields

$$n\overline{g}^{(n)} - N'_b\overline{g}^{(N'_b)} = (n - N'_b)\tau + \tfrac{1}{2}[n(n+1) - N'_b(N'_b+1)]\alpha f_Z^{-1}$$

$$\overline{g}^{(n)} = \frac{N'_b}{n}\overline{g}^{(N'_b)} + (1 - \frac{N'_b}{n})\tau + [n+1 - \frac{N'_b(N'_b+1)}{n}]\frac{\alpha}{2f_Z}. \tag{2.32}$$

Numerically, $N'_b = 5$ and at $n = 6$, $\overline{g}^{(6)} = \frac{119}{3} < 40' = t_b$.

2.5 On routing behaviour and system operations

The waiting time and the travel time considered so far are individual times as perceived by a service user who tries to minimize his travel time on a selfish behaviour. Let us define *system optimization* as the reduction of the global travel time over all passengers. This objective could be pursued using various instruments, including (i) service design by setting run time, dwelling time, operation frequency and vehicle capacity, (ii) station design by adapting the platform layout, (iii) social route choice i.e. every user obeys to a collective rationality, perhaps by following route guidance provided by a network supervisor. The latter instrument involves the evaluation of overall travel time, denoted hereafter as $G(n)$ with respect to stock size n.

When service a is available, the routing decision is whether to use it and reduce stock size, yielding system cost of $t_a + G(n-1)$, or not to use it and maintain stock size, yielding cost $G(n)$. Thus line a is socially attractive at state n if

$$t_a + G(n-1) \le G(n)\,. \tag{2.33}$$

This condition can be applied recursively to yield social attractive capacity and attractivity threshold.

Under priority queuing, $G_{\mathrm{PQ}}(n) = \sum_{i=1}^{n} \bar{g}_{\mathrm{PQ}}^{(i)}$ so social attractivity is reduced to $t_a \le \bar{g}_{\mathrm{PQ}}^{(n)}$, i.e. individual attractivity to the passenger last in stock. In other words, priority queuing is socially optimal. The reason underlying this property is the assumption that a passenger with larger rank (meaning lower priority) is able to board in a line vehicle that is unattractive to those users ranking first, by going through the stock without causing opposition.

Under mingled waiting, a myopic supervisor would evaluate $G_{\mathrm{MW}}(n)$ as $n\bar{g}_{\mathrm{MW}}^{(n)}$. Then the condition for social attractivity becomes

$$t_a \le n\bar{g}_{\mathrm{MW}}^{(n)} - (n-1)\bar{g}_{\mathrm{MW}}^{(n-1)}, \tag{2.34}$$

i.e. line run time must be less than marginal overall travel time. From the definition of $\bar{g}_{\mathrm{MW}}^{(n)}$ in (2.25), if $A(n) = A(n-1) = A$ then

$$\begin{aligned} & n\bar{g}_{\mathrm{MW}}^{(n)} - (n-1)\bar{g}_{\mathrm{MW}}^{(n-1)} \\ &= \sum_{a\in A} \frac{f_a}{f_A}[(k_{a/n} - k_{a/n-1})t_a + (n-k_{a/n})\bar{g}_{\mathrm{MW}}^{(n-k_{a/n})} - (n-1-k_{a/n-1})\bar{g}_{\mathrm{MW}}^{(n-1-k_{a/n-1})}] \end{aligned}$$

As will be shown in Section 4, $k_{a/n} - k_{a/n-1} = 1_{\{k_{a/n-1} < k_{a/n}\}} = 1_{\{n-N_a \in [1,.k_a]\}}$ if $a \in A$. Consequently,

$$n-1-k_{a/n-1} = n-1-k_{a/n} + 1_{\{k_{a/n-1}<k_{a/n}\}} = n-k_{a/n} + 1_{\{k_{a/n-1}=k_{a/n}\}}\,.$$

Furthermore, if $a \in A$ then $n-1 > N_a$ hence $1_{\{k_{a/n-1}=k_{a/n}\}}$ is reduced to $1_{\{n>N_a+k_a\}}$.

In turn,

$$\begin{aligned} & (n-k_{a/n})\bar{g}_{\mathrm{MW}}^{(n-k_{a/n})} - (n-1-k_{a/n-1})\bar{g}_{\mathrm{MW}}^{(n-1-k_{a/n-1})} \\ &= 1_{\{n>N_a+k_a\}}[(n-k)\bar{g}_{\mathrm{MW}}^{(n-k)} - (n-k-1)\bar{g}_{\mathrm{MW}}^{(n-1-k)}] \end{aligned}$$

Thus

$$\begin{aligned} & n\bar{g}_{\mathrm{MW}}^{(n)} - (n-1)\bar{g}_{\mathrm{MW}}^{(n-1)} \\ &= \sum_{a\in A} \frac{f_a}{f_A}\Big(1_{\{n-N_a\in[1,k_a]\}}t_a + 1_{\{n>N_a+k_a\}}[(n-k)\bar{g}_{\mathrm{MW}}^{(n-k)} - (n-k-1)\bar{g}_{\mathrm{MW}}^{(n-1-k)}]\Big) \end{aligned} \tag{2.35}$$

Comparing (2.35) to (2.17) and by induction on stock size, it turns out that the marginal overall travel time under mingled waiting is identical to the individual travel time under priority queuing for the last passenger in the stock – indeed a remarkable result.

This holds also if some lines not in $A(n-1)$ are included in $A(n)$, cf. Appendix §9.1.

2.6 On platform layout

Let us emphasize once again the difference between MW and PQ. Under MW, all waiting passengers have identical priority, availability and evaluation to any vehicle. Under PQ each passenger has a position in stock that amounts to a priority rank. The possibility to traverse the stock in order to board in a less attractive vehicle may be called a '*property of porosity*', which is crucial to achieve social optimum as well as selfish optimization. However priority may only be reserved *within* one stock, leading to the issue of 'Why should passenger traffic be organized into a single stock?' In practice, this depends on platform layout as implemented by the network operator. Each route service is assigned to a given dwelling slot (a point for a bus or a stretch for a train) so a stock can take place on the boarding zone associated to the slot. If the slot is assigned dynamically, as may happen in a line terminal or a large train station, then it is likely that the stock would take place in some waiting area for passengers to wait for dynamic information before and prior to a vehicle: here MW is the relevant discipline.

Let us restrict ourselves to slots and boarding zones assigned in the long run to given services. Then one boarding zone is assigned to a subset of transit services; in train transportation sharing the zone requires sharing the running track, which implies that the assigned services must be operated in a coordinated manner – they must belong to one line of railway operation.

Taking position in a given boarding zone supplies a passenger with some priority to access the associated services over those passengers who would decide dynamically (e.g. at the arrival of a vehicle) to try to board. So if the zone is populated by a stock that would saturate the next vehicle, there is no possibility for any other customer to succeed in immediate boarding. Then the user's route choice must proceed in two steps: first, the selection of a waiting place among the boarding areas; second, the choice of a service associated to that area – or alternatively of a service that would dwell at an unsaturated area.

To sum up, station layout involves fixed or dynamic assignment of services to dwelling slots. Dynamic assignment requires a specific area to wait for dynamic information. Platform layout involves one or several slots. Services that share a given slot are likely to belong to one line of operation. When several services are operated, if some of the associated areas are saturated then every customer must choose their route in two stages, first a waiting place then a service. The rest of the paper deals with one boarding zone only.

3. Priority Queuing model

Priority Queuing involves the axiomatic assumption that a customer with more priority is likely to enjoy a reduced travel time (or cost), at least on the average if not in all instances. Denoting by m the customer rank in the stock of passengers, this is stated as follows:

Assumption 1, PQ regularity: *the system is regular up to rank* n *if* $\overline{\theta}_m \leq \overline{\theta}_{m+1}$ $\forall m < n$.

3.1 Attractivity conditions and concepts

Definition 1, PQ attractivity. *Line* a *is attractive at order* m *if* $t_a \leq \overline{\theta}_m$.

PQ attractivity means that if a vehicle of line a is available, then it yields profit to the customer in rank m, either because he gets a place on board or his rank is decreased (i.e. improved).

Definition 2, The PQ attractivity threshold *of line* a, *denoted as* N_a, *is the maximum rank* $m \geq 1$ *at which line* a *is unattractive, or zero otherwise.*

By transitivity of inequality, if $\overline{\theta}_m$ is increasing up to $m+1$ then $t_a \leq \overline{\theta}_{N_a+1} \leq \ldots \leq \overline{\theta}_m \leq \overline{\theta}_{m+1}$.

Proposition 1, PQ attractivity continuation. *Assuming PQ regularity up to* n, *line* a *is attractive at all orders from* N_a+1 *to* n.

Then any customer ranking from N_a+1 to N_a+k_a would take a vehicle if line a were available.

The correct number of boarding passengers, however, is bounded by the number of such customers on the platform, yielding that:

Definition 3, The PQ attractive capacity *of line* a *with respect to rank* n *is*

$$k_{a/n} \equiv \min\{k_a, (n-k_a)^+\}. \tag{3.1}$$

3.2 Composed travel time and attractive set

Assume that the sequence $\overline{\theta}_m$ is known at least up to order $n-1$: then the attractive thresholds which are less than or equal to $n-1$ are known, since $N_a = \max\{m : \overline{\theta}_m \le t_a\}$. Put more correctly, the attractive threshold of any line a with $t_a < \overline{\theta}_{n-1}$ is known with certainty and satisfies that $N_a < n-1$, while a line with $t_a \ge \overline{\theta}_{n-1}$ may have either $N_a = n-1$ if t_a were $< \overline{\theta}_n$ or $N_a \ge n$ otherwise.

Let us associate to each line a candidate time for potential attractivity at order n.

Definition 4. The PQ composed time *of line* a *at order* n *is*

$$\hat{t}_a^{(n)} \equiv 1_{\{n \le N_a+k_a\}} t_a + 1_{\{n > N_a+k_a\}} \overline{\theta}_{n-k_a}. \tag{3.2}$$

Lemma 1. *(i) If the* $\overline{\theta}_m$ *are known up to* $n-1$ *and* $k_a \ge 1$ *then* $\hat{t}_a^{(n)}$ *is unambiguous. (ii) Assuming regularity up to* $n-1$, *it holds that* $\hat{t}_a^{(n)} \ge \hat{t}_a^{(n-1)}$.

Proof. (i) If $t_a < \overline{\theta}_{n-1}$ then $N_a < n-1$ yielding an unambiguous result. Otherwise, if $t_a \ge \overline{\theta}_{n-1}$ then $N_a \ge n-1$ hence $N_a + k_a \ge n$ since $k_a \ge 1$, so that $\hat{t}_a^{(n)} = t_a$.

(ii) If $n \le N_a + k_a$ then so is $n-1$ hence $\hat{t}_a^{(n)} = t_a = \hat{t}_a^{(n-1)}$. If $n > N_a + k_a$ then $\hat{t}_a^{(n)} = \overline{\theta}_{n-k_a}$: either $n-1 > N_a + k_a$ so that $\hat{t}_a^{(n-1)} = \overline{\theta}_{n-1-k_a} \le \overline{\theta}_{n-k_a}$ by regularity since $n - k_a \le n-1$, or $n-1 = N_a + k_a$ hence $\hat{t}_a^{(n-1)} = t_a \le \overline{\theta}_{N_a+1} = \overline{\theta}_{n-k_a} = \hat{t}_a^{(n)}$ owing to the definition of the attractivity threshold.

Definition 5, bundle cost. *At order* n *the cost of a line bundle* $B \subset Z$ *is defined as*

$$\overline{g}_B^{(n)} \equiv \frac{\alpha}{f_B} + \sum_{a \in B} \frac{f_a}{f_B} \hat{t}_a^{(n)}. \tag{3.3}$$

3.3 Attractive set and user equilibrium

Definition 6, Attractive set at order n. *This is a line bundle* $B \subset Z$ *such that*

$$t_a < \overline{g}_B^{(n)} \Rightarrow a \in B, \tag{3.4a}$$

$$t_a > \bar{g}_B^{(n)} \Rightarrow a \notin B. \tag{3.4b}$$

These conditions state the demand side of User Equilibrium (UE) in a transit system. Let us show that these correspond to a travel strategy (here a line bundle) of minimal cost.

Theorem 1, Existence of an Attractive Set at order n under PQ: *Assuming that PQ regularity holds up to $n-1$, then at order n:*

(i) There exists a line bundle $B(n) \subset Z$ that minimizes $\bar{g}^{(n)}(B)$.

(ii) $\bar{\theta}_n \equiv \bar{g}_{B(n)}^{(n)} \geq \bar{\theta}_{n-1}$.

(iii) Any bundle $B(n)$ of minimal cost is attractive.

(iv) $\forall a \notin B(n), N_a \geq n$.

(v) $\forall a \in B(n), t_a > \bar{\theta}_{n-1} \Rightarrow N_a = n-1$.

(vi) Any attractive bundle is of minimal cost.

Proof. (i) Assuming that Z is finite, then the set of bundles is finite, yielding a finite set of real values $\bar{g}^{(n)}(B)$, among which one is minimal.

(ii) From Lemma 1 point (ii), $\hat{t}_a^{(n)} \geq \hat{t}_a^{(n-1)}$ hence $\bar{g}_{B(n)}^{(n)} \geq \bar{g}_{B(n)}^{(n-1)} \geq \bar{g}_{B(n-1)}^{(n-1)}$ which defines $\bar{\theta}_{n-1}$, due to point (i) in the Theorem applied at the previous order. Thus $\bar{\theta}_n \geq \bar{\theta}_{n-1}$.

(iii) As in the uncapacitated problem of optimal travel strategy: if $t_a < \bar{\theta}_n$ and $a \notin B$ then by regularity $\hat{t}_a^{(n)}$ is less than $\bar{\theta}_n$ so $B' \equiv B \cup \{a\}$ would improve on minimal cost, which would contradict the assumption of optimality. Similarly, if $t_a > \bar{\theta}_n$ and $a \in B$ then $\hat{t}_a^{(n)} = t_a$ so $B' \equiv B - \{a\}$ would yield a cost lower than $\bar{\theta}_n$.

(iv) By (iii) and contraposition of (3.4a), $a \notin B(n) \Rightarrow\ t_a \geq \bar{\theta}_n$: thus $N_a \geq n$.

(v) By (iii) and contraposition of (3.4b), $a \in B(n) \Rightarrow t_a \leq \bar{\theta}_n$ yielding $N_a \leq n-1$. But if $t_a > \bar{\theta}_{n-1}$ then $N_a \geq n-1$: the conjunction of both conditions yields $N_a = n-1$.

(vi) It holds that $\forall a \in Z$, $t_a^{(n)} \leq \min\{\bar{\theta}_n, t_a\}$ since if $t_a > \bar{\theta}_n$ then $N_a \geq n$ hence $N_a + k_a > n$ hence $t_a^{(n)} = t_a$. Let us consider an attractive bundle A with cost $\bar{g}_A^{(n)}$: attractiveness implies that $\forall a \in Z - A$, $t_a \geq \bar{g}_A^{(n)}$ while by the optimality of $B(n)$, $\bar{g}_A^{(n)} \geq \bar{g}_{B(n)}^{(n)}$, so $t_a \geq \bar{\theta}_n$. Now let $A' \equiv \{a \in Z : t_a \geq \bar{\theta}_n\}$: it holds that $\forall a \in A'$, $t_a^{(n)} = t_a$. If $A' \neq \varnothing$ then its lines can be pooled into an option of frequency $f_{A'}$ and average run time $t_{A'} = (\sum_{a \in A'} f_a t_a)/f_{A'}$ that is less than $\bar{g}_A^{(n)}$ by the attractivity property which is maintained through convex combination. By lemma 0, if $\bar{g}_A^{(n)} > t_{A'}$ then $t_{A'} \leq \bar{g}_{A-A'}^{(n)}$ and $\bar{g}_A^{(n)} \leq \bar{g}_{A-A'}^{(n)}$: but $\bar{g}_{A-A'}^{(n)} = \bar{g}_{B(n)}^{(n)}$ because their definitions are equivalent. So $\bar{g}_A^{(n)} = \bar{g}_{B(n)}^{(n)}$ i.e. the attractive bundle has an optimal cost.

Corollary 1: Existence and uniqueness of User Equilibrium (UE) under PQ.

(i) Any attractive bundle is an optimal travel strategy for the user waiting at n.

(ii) The optimal cost is consistent with the relative capacities, so the attractive bundle is a UE state.

(iii) The optimal value $\overline{\theta}_n$ is unique and there is a unique minimal attractive set of optimal cost, defined as $\tilde{B}(n) \equiv \{a \in B(n) : t_a < \overline{\theta}_n\}$.

Proof. (i) Stems from point (vi) in Theorem 1.

(ii) stems from the definition of function $\overline{g}^{(n)}$, which implies that any vehicle of line $a \in B(n)$ will be demanded by $k_{a/n} \geq 1$, whereas point (iv) of Th.1 implies that $k_{a/n} = 0$ for $a \notin B(n)$, yielding a consistent assessment of cost by the n-th user.

(iii) involves point (i) in Th.1 and Lemma 0 – there would be no loss of optimality by outsourcing any line with $t_a = \overline{\theta}_n$ from the optimal bundle, while optimality would be lost by outsourcing any line with $t_a < \overline{\theta}_n$.

3.4 The formation of optimal bundles

Corollary 2: Continuation of regularity under PQ. *Assume that sequence $\overline{\theta}_n = g_{B(n)}^{(n)}$ is obtained incrementally: then regularity holds at any order.*

Proof. At order 1, $\hat{t}_a^{(n)} = t_a$ for all $a \in Z$ if $k_a \geq 1$. Thus $\overline{\theta}_1 = \overline{\theta}_0$ which yields regularity at order 1. Now, if regularity holds up to $n-1$, then Lemma 1 holds and so does Th.1, of which point (ii) ensures that $\overline{\theta}_n \geq \overline{\theta}_{n-1}$ i.e. regularity up to order n included. By induction, regularity holds at any order.

Corollary 3: Enlargement of optimal bundle under PQ. *The minimal optimum bundle of order m is included in the optimal bundles of any subsequent order.*

Proof. If $a \in \tilde{B}(m)$ minimal then $t_a < \overline{\theta}_m$ hence $t_a < \overline{\theta}_n$ $\forall n \geq m$: then point (iii) in Th.1 at order n implies that $a \in B(n)$.

The recursive formation of the minimum cost and the associated optimal bundle amounts to a recursive algorithm to solve for UE up to any order. Based on Corollary 3, from order $n-1$ an efficient implementation would be to test $A_0^{(n)} \equiv B(n-1)$ as a candidate optimal bundle at order n, yielding cost $\overline{g}_0^{(n)}$: having ranked the lines a in order of increasing t_a, then those lines not in $A_i^{(n)}$ but with $t_a < \overline{g}_i^{(n)}$ should be included in $A_{i+1}^{(n)}$: this should end up when there remains no line or the next line satisfies $t_a > \overline{g}_i^{(n)} \equiv \overline{g}^{(n)}(A_i^{(n)})$, yielding $B(n) \equiv A_i^{(n)}$.

4. Mingled Waiting model

In this Section, MW is addressed in much the same manner as PQ in Section 3, except for one significant peculiarity: under MW the composed time of a line may be less than the line run time even if the attractive capacity is strictly positive at that order. This obliges us to state more elaborate conditions for UE.

4.1 Basic concepts

Recall that the main state variable under MW is the stock size, denoted by n.

Definition 7, The recourse option *to line a under stock size n is to remain on platform rather than to board in a vehicle of line a. Let $\overline{\theta}_n^{-a}$ denote its average travel time (or cost).*

The sequences $(\overline{\theta}_n^{-a})_{n\geq 0}$ for $a \in Z$ summarize the waiting strategy of a user.

Definition 8, The relative capacity *of line a with respect to the recourse option $(\overline{\theta}_m^{-a})_m$ and stock size n is the integer number $k_{a/n}$ that solves the following system:*

$$\max i \textit{ such that} \tag{4.1a}$$

$$0 \leq i \leq k_a, \tag{4.1b}$$

$$i \leq n, \tag{4.1c}$$

$$t_a \leq \overline{\theta}_{n-i+1}^{-a} \textit{ if } i \geq 1. \tag{4.1d}$$

If sequence $(\overline{\theta}_m^{-a})_m$ increases up to n, then the solution i of (4.1) is limited by one of its linear constraints. If $i = \min\{k_a, n\}$ then (4.1d) holds at i and potentially beyond i. If $i < \min\{k_a, n\}$ then (4.1d) cannot hold at $i+1$ i.e. $t_a > \overline{\theta}_{n-i}^{-a}$.

Definition 9, The MW attractivity threshold *of line a with respect to the recourse option $(\overline{\theta}_m^{-a})_m$ is*

$$N_a \equiv \max\{m \geq 0 : t_a \leq \overline{\theta}_{m+1}^{-a}\}. \tag{4.2}$$

Lemma 2. *If sequence $(\theta_m^{-a})_m$ increases with m up to n then the relative capacity and the attractivity threshold of line a are linked by*

$$k_{a/n} = \min\{k_a, (n-N_a)^+\}. \tag{4.3}$$

Proof. Let $k = k_{a/n}$. From its definition in (4.1b-c), $k \geq 0$ and $k \leq \min\{k_a, n\}$. If $n > 0$ and $k = 0$ then (4.1d) does not hold at $i = 1$, which requires that $n \leq N_a$ since the sequence $\overline{\theta}_m^{-a}$ increases : then $(n - N_a)^+ = 0$ which satisfies (4.3). If $k > 0$ then in the definitional program (4.1d) is equivalent to $i \leq (n - N_a)^+$, hence the maximal solution is $\min\{k_a, (n-N_a)^+\}$ i.e. (4.3).

Thus relative capacity is also an attractive capacity.

4.2 Axiomatic regularity and its consequences

Definition 10, *Under MW and stock n, travel time to the destination is a random variable, denoted θ_n, with average value denoted as $\overline{\theta}_n$.*

Definition 11: MW regularity. *The system is regular up to order n if, $\forall m < n$,*

$$\overline{\theta}_m \leq \overline{\theta}_{m+1}, \tag{4.4a}$$

$$\overline{\theta}_m^{-a} \leq \overline{\theta}_{m+1}^{-a}, \ \forall a \in Z, \tag{4.4b}$$

$$\overline{\theta}_m \leq \overline{\theta}_m^{-a}, \ \forall a \in Z. \tag{4.4c}$$

Definition 12: MW composed time. *To line a with respect to stock size n, the composed time is the random variable that mixes t_a with $\theta_{n-k_{a/n}}$ in proportion of $k_{a/n}/n$ and $1-k_{a/n}/n$, respectively. At $n=0$ by convention $\frac{1}{n}k_{a/n}=1_{\{t_a\le\bar{\theta}_0\}}$. The expected composed time is*

$$\bar{t}_a^{(n)} \equiv \frac{k_{a/n}}{n}t_a + (1-\frac{k_{a/n}}{n})\bar{\theta}_{n-k_{a/n}}. \quad (4.5)$$

Lemma 3, origin of a composed time. *Assuming that (i) sequence $\bar{\theta}_n$ increases with m up to n, (ii) $\bar{\theta}_m \le \bar{\theta}_m^{-a}$ and (iii) $N_a \ge 1$, then $\bar{t}_a^{(m)} < t_a$ at least up to N_a+1.*

Proof. For $n \le N_a$, $k_{a/n}=0$ hence $\bar{t}_a^{(n)} = \bar{\theta}_n$ which is less than $\bar{\theta}_n^{-a}$ by (ii), hence $<t_a$ since $n \le N_a$. At $n = N_a+1$, $\bar{t}_a^{(n)}$ is a convex combination with strictly positive coefficients of t_a and $\bar{\theta}_{N_a}$ which is $<t_a$, which implies that the outcome is $<t_a$.

This property distinguishes the capacitated model from the uncapacitated one, in which any attractive line checks that $t_a \le \bar{g}_A$ as stated in Section 2. This inequality holds for a capacitated line only if it is attractive from the origin i.e. $N_a=0$.

Lemma 4, development of the expected composed time. *Assuming that (i) regularity holds up to n and (ii) $t_a \le \bar{\theta}_m^{-a}$ for $m \le n$. Then $\bar{t}_a^{(i)}$ increases with $i \in \{m, m+1, ...n, n+1\}$.*

Proof. By assumption $N_a < m$ hence $N_a < i$. By (4.3) the sequence $k_{a/i}$ increases with i. If $i - N_a < k_a$ then $k_{a/i+1} = k_{a/i}+1$ (which holds also at $i=n$ even if $\bar{\theta}_{n+1}^{-a}$ is not yet known) hence $i+1-k_{a/i+1} = i-k_{a/i} = N_a$ so that, letting $k=k_{a/i}$,

$$\begin{aligned}\bar{t}_a^{(i+1)} - \bar{t}_a^{(i)} &= t_a(\tfrac{1+k}{1+i}-\tfrac{k}{i}) + \bar{\theta}_{i-k}(i-k)(\tfrac{1}{i+1}-\tfrac{1}{i}) \\ &= \tfrac{N_a}{i(1+i)}(t_a - \bar{\theta}_{N_a}) \\ &\ge 0\end{aligned}$$

since if $N_a=0$ then the right-hand side is zero whereas if $N_a>0$ then $t_a > \bar{\theta}_{N_a}^{-a} \ge \bar{\theta}_{N_a}$ from the definition of N_a and assumption (ii).

If $i - N_a \ge k_a$ then $k_{a/i+1} = k_{a/i} = k_a$ denoted as k, yielding that

$$\begin{aligned}\bar{t}_a^{(i+1)} - \bar{t}_a^{(i)} &= \tfrac{-k}{i(1+i)}t_a + \bar{\theta}_{i+1-k}\,\tfrac{i+1-k}{i+1} - \bar{\theta}_{i-k}\,\tfrac{i-k}{i} \\ &= -\tfrac{k}{i(1+i)}(t_a - \bar{\theta}_{i+1-k}) + \bar{\theta}_{i+1-k} - \bar{\theta}_{i-k}\end{aligned}$$

The outcome is ≥ 0 since by assumption $\bar{\theta}_{i+1-k} - \bar{\theta}_{i-k} \ge 0$ and because one out of the two following conditions holds:

- either $t_a \le \bar{\theta}_{i-k}$ so, as $\bar{\theta}_{i-k} \le \bar{\theta}_{i+1-k}$, by transitivity $t_a \le \bar{\theta}_{i-k+1}$ which yields the claimed outcome,

- or $t_a > \overline{\theta}_{i-k}$ which implies that $\overline{\theta}_{i+1-k} - \overline{\theta}_{i-k} \geq \overline{\theta}_{i+1-k} - t_a$. Then, as $i > k$, $i(i+1)(\overline{\theta}_{i+1-k} - \overline{\theta}_{i-k}) + k(\overline{\theta}_{i+1-k} - t_a) \geq (i^2 + i - k)(\overline{\theta}_{i+1-k} - \overline{\theta}_{i-k}) \geq 0$, which yields that $\bar{t}_a^{(i+1)} - \bar{t}_a^{(i)} \geq 0$.

4.3 MW attractivity and user equilibrium

Definition 13: Relative attractivity. *Line a is attractive with respect to recourse option $(\overline{\theta}_m^{-a})_m$ and stock size n iff*

$$t_a \leq \overline{\theta}_n^{-a}. \tag{4.6}$$

This condition is microeconomic as it characterizes the user's individual choice behaviour: were line a immediately available, the user would compare its run cost t_a to the recourse cost of waiting for another service, $\overline{\theta}_n^{-a}$.

By the transitivity of inequality, if the recourse costs $\overline{\theta}_m^{-a}$ increase up to n then if line a is attractive at m, so it remains at any order $i \in \{m, m+1..n\}$.

Definition 14: Average bundle cost. *At order n the expected cost of a line bundle $B \subset Z$ is, under MW,*

$$\overline{g}_B^{(n)} \equiv \frac{\alpha}{f_B} + \sum_{a \in B} \frac{f_a}{f_B} \bar{t}_a^{(n)}. \tag{4.7}$$

Definition 15: MW Attractive bundle. *At order n a line bundle $B \subset Z$ is attractive iff, denoting by $B^* \equiv \arg\max\{t_b : b \in B\}$:*

$$\forall a \in B,\ t_a \leq \max\{\overline{\theta}_{n-1}^{-a}, \overline{g}_{B-B^*}^{(n)}\}, \tag{4.8a}$$

$$\forall a \in Z - B,\ t_a \geq \overline{g}_B^{(n)}. \tag{4.8b}$$

Condition (4.8a) provides a precise definition for $\overline{\theta}_n^{-a}$ in (4.6) on the basis of, first, the route choice opportunities at the previous order, second, a recourse option $B - B^*$ that is more elaborate than $B - a$ and involves some transitivity. When there are several lines in the bundle, a recourse cost of the form $\overline{g}_{B-a}^{(n)}$ might not be greater than t_a would the bundle include a line other than a and that would have a little larger run time but a much higher frequency.

These are theoretical conditions for user equilibrium in transit traffic under MW: (a) the cost t_a of an attractive option that is immediately available is less than the highest recourse cost among attractive options, (b) an unattractive option even if it is immediately available has a larger cost than that of bundle B.

The set of conditions (4.8) involves three significant differences from the classical Wardrop conditions for assignment to a private mode of transportation. First, the fragmentary availability of transit services requires to distinguish between line cost when available and travel cost. Second, the distinction between $\overline{g}_{B-B^*}^{(n)}$ and $\overline{g}_B^{(n)}$ is necessary to state the influence of capacity constraints on attractivity under MW – contrary to both uncapacitated or capacitated under PQ. Third, the recursive formation of composed cost hence of bundle cost internalizes the capacity constraints, which differs from the formulations of capacitated User Equilibrium for private transportation that involves dual variables (e.g. Larsson and Patriksson, 1994).

4.4 The recursive structure of attractive bundles

At order $n = 1$, the composed cost of any capacitated line with $k_a \geq 1$ is reduced to t_a, yielding user equilibrium (UE) conditions (4.8) that amount to the uncapacitated problem of attractivity. The $\overline{\theta}_1^{-a}$ must be set to $\overline{g}_{B(1)-a}^{(1)}$.

This provides the origin for the progressive determination of UE at any subsequent order.

Theorem 2: Existence and uniqueness of UE at order n under MW. *Assume that $t_a \leq \overline{\theta}_{n-1}^{-a}$ $\forall a \in B(n-1)$. At order n let us associate to any bundle $B \subset Z$ a companion bundle $C_B^{(n)} \equiv \{b \in Z - B : t_b < \overline{g}_B^{(n)}\}$.*

(i) Any attractive bundle satisfies that $C_B^{(n)} = \varnothing$.

(ii) If $B(n-1) \neq \varnothing$ then there exists an attractive bundle $B(n)$ at order n.

(iii) $B(n)$ contains $B(n-1)$: the residual lines $b \in B(n) - B(n-1)$ have $N_b = n-1$, while those in $Z - B(n)$ have $N_b \geq n$.

(iv) Let $\overline{\theta}_n \equiv \overline{g}_{B(n)}^{(n)}$, $\overline{\theta}_n^{-a} \equiv \overline{\theta}_n$ $\forall a \in Z - B(n)$ and $\overline{\theta}_n^{-a} \equiv \max\{\overline{\theta}_{n-1}^{-a}, \overline{g}_{B(n)-B(n)^*}^{(n)}\}$ $\forall a \in B(n)$: then $\overline{\theta}_n \geq \overline{\theta}_{n-1}$, $\overline{\theta}_n^{-a} \geq \overline{\theta}_{n-1}^{-a}$ and $\overline{\theta}_n \leq \overline{\theta}_n^{-a}$.

(v) Bundle $B(n)$ so determined is unique: any other attractive bundle would be the union of $B(n)$ and other lines b with $t_b = \overline{g}_{B(n)}^{(n)}$, yielding identical cost.

Proof. (i) If B is attractive then (4.8b) holds so any $z \in Z - B$ has $t_z \geq \overline{g}_B^{(n)}$, thus making $C_B^{(n)}$ an empty set.

(ii) Lemma 4 ensures that $\overline{t}_a^{(n)} \geq \overline{t}_a^{(n-1)}$ $\forall a \in Z$ so that for any bundle B, $\overline{g}_B^{(n)} \geq \overline{g}_B^{(n-1)}$. Thus (4.8a) is satisfied at order n for the lines of maximal run time in $B(n-1)$. Now, if $C_{B(n-1)}^{(n)}$ is empty then (4.8b) holds, too, making $B(n-1)$ an attractive set at order n.

If $C_{B(n-1)}^{(n)} \neq \varnothing$ then apply the following algorithm to augment $B(n-1)$ into bundle B. While $C_B^{(n)} \neq \varnothing$, take b in it with minimum t_b: by definition $t_b < \overline{g}_B^{(n)}$, which ensures that in turn $\overline{t}_b^{(n)} \equiv \frac{1}{n} t_b + (1 - \frac{1}{n}) \overline{\theta}_{n-1}$ is $< \overline{g}_B^{(n)}$. Consider now $B' \equiv B \cup \{b\}$: as $t_b < \overline{g}_B^{(n)}$ and $B = B' - b$, (4.8a) holds for B' since $b \in \arg\max\{t_a : a \in B'\}$. Furthermore, by Lemma 0 it holds that $\overline{g}_{B'}^{(n)} < \overline{g}_B^{(n)}$. As $b \in C_{B(n-1)}^{(n)}$ then $t_b \geq \overline{\theta}_{n-1}$ and $\overline{t}_b^{(n)} \geq \overline{\theta}_{n-1}$ in turn, yielding $\overline{g}_{B'}^{(n)} \geq \overline{\theta}_{n-1}$ by convex combination of two terms each one larger than $\overline{\theta}_{n-1}$. The companion set $C_{B'}^{(n)}$ must be smaller than $C_B^{(n)}$ since the limiting cost is decreased. By replacing B with B', the process must terminate in at most $\mathrm{Card}(C_{B(n-1)}^{(n)})$ steps, yielding in the end a final B' that satisfies (4.8b) as well as (4.8a): this is kept as $B(n)$ and its line of maximum run time as $B^*(n)$.

(iii) From (ii), $B(n-1) \subset B(n)$. If $b \notin B(n-1)$ then $t_b \geq \overline{\theta}_{n-1}$ so that $N_b \geq n-1$. If $b \in B(n)$ then $t_b < \overline{g}_{B-B^*}^{(n)}$ so that $N_b < n$: at the intersection $b \in B(n) - B(n-1)$, $N_b = n-1$.

(iv) $\overline{g}_B^{(n)} \geq \overline{\theta}_{n-1}$ has been shown to hold throughout the building process, yielding in the end $\overline{g}_{B(n)}^{(n)} \geq \overline{\theta}_{n-1}$.

If $a \in Z - B(n)$ then $\overline{\theta}_{n-1}^{-a} = \overline{\theta}_{n-1}$ and $\overline{\theta}_n^{-a} = \overline{\theta}_n$ so that $\overline{\theta}_n^{-a} \geq \overline{\theta}_{n-1}^{-a}$ and $\overline{\theta}_n^{-a} \geq \overline{\theta}_n$.

If $a \in B(n)$, then either $a \in B(n-1)$ in which case the basic assumption $t_a \leq \overline{\theta}_{n-1}^{-a}$ applies, or $a \in B(n) - B(n-1)$ in which case $t_a \leq t_b$ with $b \in B^*(n)$ so $t_b \leq \overline{g}_{B(n)-B^*}^{(n)}$ yielding $t_a \leq \overline{g}_{B(n)-B^*}^{(n)}$: in both cases $t_a \leq \overline{\theta}_n^{-a}$ from its definition, which also implies that $\overline{\theta}_n^{-a} \geq \overline{\theta}_{n-1}^{-a}$ and $\overline{\theta}_n^{-a} \geq \overline{\theta}_n$ since $\overline{g}_{B(n)-B^*}^{(n)} \geq \overline{\theta}_n$.

(v) At every order $n \geq 2$, the process of line inclusion into current bundle cannot stop sooner unless it begins with a strictly smaller set $B(n-1)$. By induction, this would compel $B(1)$ to be smaller than it is, which cannot hold since there exists a minimum solution to the uncapacitated problem that is unique except for degenerate cases with $t_b = \overline{g}_{B(1)}^{(1)}$. So the sequence $\overline{\theta}_n$ is unique, and any attractive bundle must include the minimum attractive bundle plus eventually some lines that are degenerate at the current order.

4.5 An incremental user equilibrium algorithm

To determine an attractive bundle i.e. a state of UE at any order n requires to determine attractive bundles at all previous orders so as to obtain the composed costs and the attractivity thresholds. Here is a streamlined algorithm:

> *Origin*. Solve the uncapacitated UE, yielding $B(1) = \{a \in Z : N_a = 0\}$, $\overline{\theta}_1$ and $\overline{\theta}_1^{-a}$ $\forall a \in Z$. Let $m := 2$.
>
> *Progression*. Based on $(\overline{\theta}_\ell)_{\ell<m}$, evaluate the composed costs $t_a^{(m)}$ for $a \in B(m-1)$ and let $B := B(m-1)$.
>
> *Inclusion process*. While $C_B^{(m)} \neq \varnothing$ do: select $b \in \arg\min\{t_b : b \in C_B^{(m)}\}$, let $B' := B \cup \{b\}$, $\overline{t}_b^{(m)} := \frac{1}{m} t_b + \frac{m-1}{m} \overline{\theta}_{m-1}$, $f_{B'} := f_B + f_b$, $\overline{g}_{B'}^{(m)} := [f_B \overline{g}_B^{(m)} + f_b \overline{t}_b^{(m)}] / f_{B'}$, then replace B by B'.
>
> *Termination test*. Let $B(m) := B'$, $\overline{\theta}_m := \overline{g}_{B'}^{(m)}$ and $\overline{\theta}_m := \overline{g}_{B'}^{(m)}$. If $m = n$ then terminate else let $m := m+1$ and go to Progression.

The space complexity of the algorithm is $O(n.|Z|)$. So is its time complexity since each order requires to evaluate the composed costs on all of the presently attractive lines; the inclusion process involves less operations since any line will become attractive only once. Of course it is convenient to deal with set Z as a list of lines ordered by increasing run time.

5. From traffic theory to network assignment

Our theory of passenger waiting and route choice at a transit station amounts to a fundamental traffic diagram for a bundle of transit lines, i.e. a relationship between three state variables of, respectively, concentration, flow and speed or travel time (§ 5.1). Thus travel time can be modelled as a function of either stock size or exit flow (§ 5.2). The model of route choice yields formulae to split the passenger flow between the transit lines, with remarkable limit properties for large stock sizes (§ 5.3). Also provided are approximate formulae for travel time and wait time under the assumption of continuous and constant stock and flow (§ 5.4). Lastly, some references are made to previous models of capacitated transit assignment (§ 5.5).

5.1 A fundamental traffic diagram for a line bundle

Since the seminal paper of Greenshields (1935), the fundamental diagram of traffic has been a cornerstone in the analysis of car traffic along a roadway section (e.g. HCM, 2010). It involves three state variables that are (i) the vehicle concentration or density (in veh/km by traffic lane), (ii) the average flow speed (in km/h) or equivalently the travel time and (iii) the vehicle flow (in veh/h). The diagram relates one variable to another, notably speed versus density or flow versus density. Furthermore, under the stationary regimes of traffic the flow is equal to the product of speed and density.

For a line bundle, i.e. a set of transit services available from a station platform (which makes the tail node of each service as modelled as a network link), the three state variables are, respectively: (i) the size of passenger stock, n; (ii) the average individual travel time to destination, $\overline{\theta}_n$; (iii) the exit flow that leaves the platform, $\bar{x}_Z$. Using the previous notation, the relationship between $\overline{\theta}_n$ and n has been indicated for waiting discipline either priority queuing or mingling. The relationship between the flow and stock size is as follows:

$$\bar{x}_Z(n) = \sum_{z \in Z} f_z \, k_{z/n} \,, \tag{5.1}$$

which is based on the relative capacities. Recalling that $k_{z/n} = \min\{k_z, (n - N_z)^+\}$, it is an increasing function of n so $\bar{x}_Z(n)$ increases with n, too.

Indeed, relationship (5.1) is the transit counterpart of the traffic linkage between car flow, speed and density in roadway traffic: stock size and density are analogous, while line frequency plays a role similar to flow speed.

5.2 Travel time function for a line bundle

As the average travel time also increases with n, there exists an increasing relationship between the travel time and the exit flow: let us denote it by $\theta = \mathrm{T}_Z(x)$, in which

$$\mathrm{T}_Z \equiv \overline{\theta} \circ \mathrm{x}_Z^{-1}. \tag{5.2}$$

Similar properties do not hold in general for the average waiting time when there are several lines, because if a given line becomes attractive then its inclusion in the attractive bundle reduces the wait time. When the bundle contains one line only, wait time is an increasing function of stock size (or exit flow).

Of course, our model of transit bundle is limited to one destination and homogeneous passengers. Realistic application requires dealing with important network features such as flow loading along a service line, which relates the capacity provided at a given station to the entry and exit flows at the upstream stations. However our model demonstrates that passenger stock is an important state variable, whatever the traffic regime dynamic or static i.e. stationary. A simple problem of static assignment would be to impose an origin-destination volume of customers, say q during a reference period H, to the line bundle: assuming stationarity, the equality between entry and exit flow and the condition $q = \mathrm{x}_Z(n)$ would yield n:

$$n = \mathrm{x}_Z^{-1}(q/H), \tag{5.3a}$$

$$\mathrm{T}_Z(q/H) \equiv \overline{\theta}_n . \tag{5.3b}$$

Thus the line bundle may be used as a store-and-forward element in a network problem.

5.3 Flow split formulae

Given the passenger stock, the service flow that is assigned to line a is simply

$$\overline{x}_a(n) = f_a\, k_{a/n} . \tag{5.4}$$

which amounts to attractive capacity at the service level.

Then the flow share of service a is

$$\eta_a(n) \equiv \frac{\overline{x}_a(n)}{\overline{x}_Z(n)} = \frac{f_a\, k_{a/n}}{\sum_{z\in Z} f_z\, k_{z/n}} . \tag{5.5}$$

The flow share of a given service is the ratio of its attractive capacity to the overall attractive capacity of the line bundle. Assuming saturated capacities i.e. $k_{a/n} = k_a \quad \forall a \in Z$, then

$$\eta_a = \frac{f_a\, k_a}{\sum_{z\in Z} f_z\, k_z}, \tag{5.6}$$

i.e. a ratio of supplied capacity which no longer depends on n.

Formulae (5.5) and (5.6) may be compared to that of the uncapacitated model, $p_a = f_a / \sum_{z\in A} f_z$. In the capacitated model, line frequency is multiplied by vehicle attractive capacity prior to line combination. Thus the competition of service options is summarized by the main state variable, n, which itself would stem from the travel time relationship between $\overline{\theta}_z$ as T_Z and an exogenous traffic load q as $H.\overline{\mathrm{x}}_Z$.

5.4 Continuous approximation for travel time and wait time

So far we have used recursive formulae for travel time and wait time. Let us search for some more straightforward formulas by setting ad hoc assumptions.

Under MW let us assume that attractive bundles do not depend on stock size i.e. $N_z = 0$ $\forall z \in Z$ so that $B = Z$. Then, letting $k'_a \equiv k_a - (n - k_a)^+$,

$$\overline{\theta}_n = \frac{1}{f_B}[\alpha + \textstyle\sum_{a\in B} f_a \dfrac{k'_a t_a + (n-k_a)^+\, \overline{\theta}_{n-k_a}}{n}],$$

So that, letting $f'_a \equiv f_a / f_B$ and $(f'.k'.t)_B \equiv \sum_{a\in B} f'_a\, k'_a\, t_a$,

$$n\overline{\theta}_n - \textstyle\sum_{a\in B} f'_a (n-k_a)^+\, \overline{\theta}_{n-k_a} = \dfrac{\alpha n}{f_B} + (f'.k'.t)_B .$$

Now, neglecting the potential discrepancy between $(n-k_a)^+$ and $(n-k_a)$ on one hand and between k_a and k'_a on the other, let us search for a quadratic expression $\xi+\beta n+\gamma n^2$ of $n\overline{\theta}_n$: it must hold that

$$\xi+\beta n+\gamma n^2-\sum_{a\in B} f'_a[\xi+\beta(n-k_a)+\gamma(n-k_a)^2]=\frac{\alpha n}{f_B}+(f'.k.t)_B\,,$$

which yields that

$$\gamma=\frac{\alpha}{2}\frac{1}{(f.k)_B}\,,$$

$$\beta=\frac{\alpha}{2}\frac{(f.k^2)_B}{(f.k)_B^2}+\frac{(f.k.t)_B}{(f.k)_B}\,.$$

Denoting $\kappa_B\equiv(f.k)_B$ and $\eta_a\equiv f_a k_a/\kappa_B$, then

$$\overline{\theta}_n^{\mathrm{MW}}\approx\frac{\alpha}{2\kappa_B}[n+(\eta.k)_B]+(\pi.t)_B\,. \tag{5.7}$$

The approximate average wait time stems from (5.7) after replacing t_a by 0 and reducing generalized time into physical time through division by α, yielding that

$$\overline{w}_n^{\mathrm{MW}}\approx\frac{1}{2\kappa_B}[n+(\eta.k)_B]\,. \tag{5.8}$$

Under priority queuing, consider $G^{(n)}\equiv\sum_{i=1}^{n}\overline{g}^{(i)}$ and make the same ad hoc assumptions that $N_a=0$ and $B=Z$: then

$$\begin{aligned}\frac{G^{(n)}}{n}&=\alpha.\overline{w}_B+\sum_{a\in B} f'_a\Big[\frac{k_a}{n}t_a+\frac{1}{n}\sum_{i=k_a+1}^{n}\overline{g}^{(i-k_{a/i})}\Big]\\&=\alpha.\overline{w}_B+\sum_{a\in B} f'_a\Big[\frac{k_a}{n}t_a+\frac{n-k_a}{n}\frac{G^{(n-k_a)}}{n-k_a}\Big]\end{aligned},$$

This formula can be identified with that for $\overline{\theta}_{\mathrm{MW}}^{(n)}$, yielding the approximation

$$G^{(n)}=\beta n+\gamma n^2\,. \tag{5.9}$$

As $\overline{g}^{(n)}=G^{(n)}-G^{(n-1)}$, the last formula leads to

$$\overline{g}_{\mathrm{PQ}}^{(n)}=\frac{\alpha}{(f.k)_B}n+\frac{\alpha}{2}\Big[\frac{(f.k^2)_B}{(f.k)_B^2}-\frac{1}{(f.k)_B}\Big]+\frac{(f.k.t)_B}{(f.k)_B}\,. \tag{5.10}$$

The average wait time is derived from that by replacing t_a with zero and dividing by α, yielding

$$\overline{w}_n^{\mathrm{PQ}}\approx\frac{n}{\kappa_B}+\frac{1}{2}\Big[\frac{(f.k^2)_B}{(f.k)_B^2}-\frac{1}{(f.k)_B}\Big]\,. \tag{5.11}$$

As in the binary instance addressed in Section 2, it is found here that mingled passengers estimate half the wait time that would be estimated by a passenger queued with rank equal to the stock size.

The replacement of $(n-k_a)^+$ by $(n-k_a)$ makes the formulas crude approximations. Little's law enables us to equate two statements of the total wait time experienced during a reference period H: first, that spent on platform i.e. $H.\bar{n}$, second, that wait time spent by the exited customers i.e. $H.\kappa_B.\bar{w}$: so a more robust estimation would be

$$\bar{w}^{(n)} \approx \bar{n} / \kappa_B\,, \tag{5.12}$$

which is based only on the assumption (approximation) that n remains about constant. Up to a constant term that would vanish should all vehicles have the same capacity, (5.11) and (5.12) are identical, whereas (5.8) under MW differs from them by a halving factor that could be interpreted as some optimism bias among mingled passengers.

However, as coefficient α depicts the discomfort of wait relative to in-vehicle time, another interpretation could be to distinguish α_{MW} from α_{PQ}. Then, setting $\alpha_{\mathrm{MW}} = \frac{1}{2}\alpha_{\mathrm{PQ}}$ would make the approximation formulae identical, meaning that exogenous conditions (q,θ) would yield an identical stock size whatever the waiting discipline. In an application there is no prior expectation that $\bar{n}_{\mathrm{MW}}$ and $\bar{n}_{\mathrm{PQ}}$ would be identical in response to factors (q,θ), so factors α should be calibrated to the case on the basis of data about not only time and flow but also stock size. Indeed, stock size also determines platform crowding and some specific phenomena such as the discomfort of waiting there and the processes of alighting from and boarding in the service vehicles.

5.5 Comparison to previous transit assignment models

Three main approaches have been taken to model vehicle capacity in transit assignment, namely Effective frequency, Failure to board and User preference set.

In the effective frequency model (De Cea and Fernandez, 1993, Cominetti and Correa, 2001, Cepeda et al, 2006), at any station along a transit line the line wait time increases with respect to the passenger through flow, the alighting flow and the boarding flow on the basis of a function that is a mathematical artefact; then the line frequency is derived from the wait time to be combined at that station in the classical way. This model represents neither stock size nor waiting discipline; the evaluation of wait time and line flow share is artificial.

In the Failure-to-board model (Kurauchi et al, 2003, Shimamoto et al, 2005), stock size is not represented explicitly but the waiting discipline is considered. Under MW the wait time stems from the probability of failure to board in a given vehicle, based on the residual capacity and the number of candidate riders. The attractivity of a line is determined by the sum of its expected wait time (by taking into account the random number of vehicle arrivals until success to board) and travel time. The flow share between attractive lines stems from their nominal frequency, before the boarding flow is truncated at the line residual capacity.

In the User-preference-set model (Hamdouch et al, 2004), passengers wait in a queue to board in a line vehicle or divert to another line. The wait time is not identified within the travel time: for this reason attractivity is defined in a specific way, based on an ordered list of 'travel links from current node', called the user preference set at the station node. Only the entry passenger flow is considered – not the stock size. The share of flow between alternative attractive lines is proportional to their supplied capacity.

More information about these models is provided by Leurent and Askoura (2010). It seems that our model of Passenger Stock and Attractivity Threshold (PSAT) addresses explicitly the widest set of features among the static models of transit assignment, within its own outreach.

6. Markovian model

Let us now state the PSAT model in the framework of queuing theory. Indeed, the passengers that arrive at the station platform make up a flow of customer with arrival rate λ, whereas service is delivered along transit lines of given capacity k_a that arrive at rate f_a. The state variable is the number of customers waiting in station, n. In the 'bulk service' model of queuing theory (e.g. Kleinrock, 1975) it is assumed that all capacity available in a vehicle will be attractive to any passenger waiting at the station: this is the PSAT model with null thresholds. Here the general PSAT model is cast into a Markovian framework: the waiting discipline is embedded through the threshold values, which stem from microeconomic behaviours as stated previously.

Two previous works on transit assignment are noteworthy in this context. Firstly, Chriqui and Robillard (1975) addressed the uncapacitated problem in which only attractive lines are used and their service by a vehicle empties the passenger stock whatever its size. Then the resulting stock model is very simple: the stationary probability is distributed Poisson with parameter the combined frequency of attractive lines. Secondly, Cominetti and Correa (2001) stated the transition equations for a bulk model restricted to attractive lines: but they did not identify attractivity thresholds.

Here the PSAT model is expressed as a state-transition model (§ 6.1). Then stochastic equilibrium is formulated, with explicit solution for binary models with two lines of vehicle capacity unity or infinity (§ 6.2). Macroscopic properties are derived for stock size, yielding the average wait time, the line flows and the travel time (§ 6.3). Lastly, some numerical illustration is provided (§ 6.4).

6.1 State-transition model

Assume that (i) the process of customer arrivals is Markov with rate λ, (ii) each transit line is serviced by a process of vehicle arrivals that is Markov with rate f_a and (iii) all processes are independent. Then the station as a queuing system has one state variable only, the size of the passenger stock denoted here as X. At a given instant h, the state variable has a given value $X(h) = n$ that is a nonnegative integer. Each such value n is a system state, from which a transition may occur as follows:

- To state $n+1$ with rate λ, meaning the arrival of one more customer.
- For each line $a \in Z$, to state $n - k_{a/n}$ with rate f_a meaning the arrival of a service vehicle that is used by a number $k_{a/n}$ of customers.

Any other transition between distinct states has null rate. So the Markov chain has the following infinitesimal generator $\mathbf{Q} = [q_{n,m} : n, m \geq 0]$:

$$q_{n,n+1} = \lambda \tag{6.1a}$$

$$q_{n,n+m} = 0 \text{ for } m > 1 \tag{6.1b}$$

$$q_{n,i} = \sum_{a \in Z : k_{a/n} = n-i} f_a \text{ for } i < n \tag{6.1c}$$

$$q_{n,n} = -\lambda - \sum_{a \in Z: k_{a/n} = n-i} f_a \text{ for } n > 0. \tag{6.1d}$$

Fig. 1 (resp. 2) depicts the state-transition graph of two lines a and b with $N_a = 0$ and $\nu \equiv N_b > 0$ with $(k_a, k_b) = (1,1)$ (resp. $(k_a, k_b) = (1, \infty)$).

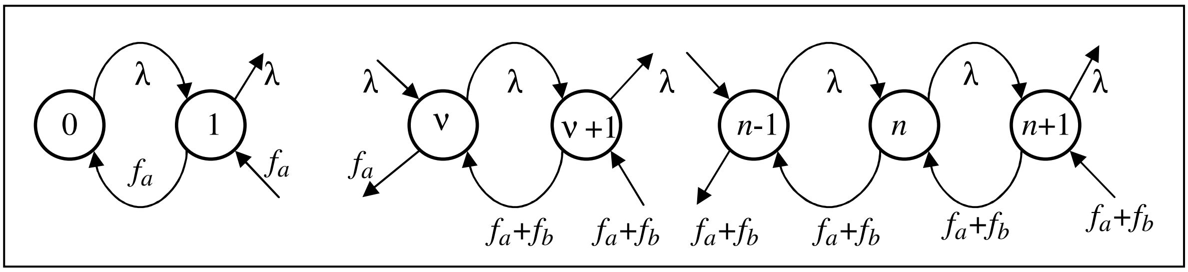

Fig. 1. State-transition graph of binary model with $(k_a, k_b) = (1,1)$.

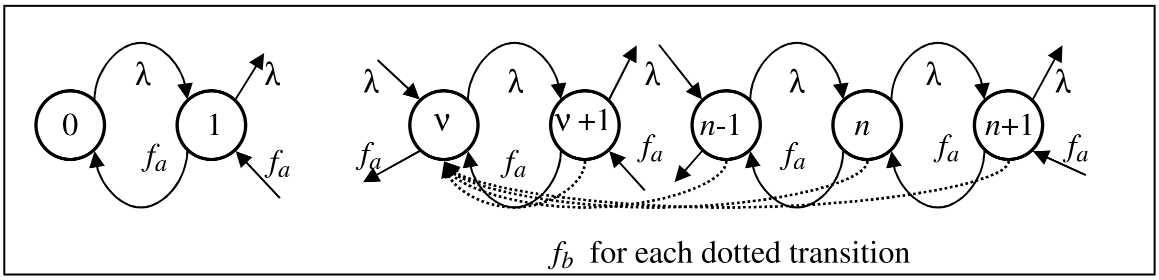

Fig. 2. State-transition graph of binary model with $(k_a, k_b) = (1, \infty)$.

6.2 Stochastic equilibrium

The system is in stochastic equilibrium (i.e. under stationary regime) if at any instant the probability of being in a given state stems from the stationary distribution $\boldsymbol{\pi} = [\pi_n]_{n \geq 0}$ that satisfies the conservation (or balance) of probability flow for each state. In other words, for each system state the exports and imports of probability flow are balanced:

$$\pi_n \sum_{m \neq n} q_{n,m} = \sum_{m \neq n} \pi_m q_{m,n} . \tag{6.2}$$

Owing to the definition of q_{nn}, this amounts to

$$\sum_m \pi_m q_{m,n} = 0 .$$

In matrix form, denoting by $\mathbf{1} = [1]_{n \geq 0}$, the conservation equation that defines the stationary distribution of probability is:

$$\boldsymbol{\pi}.\mathbf{Q} = 0 \text{, subject to } \boldsymbol{\pi} \geq \mathbf{0} \text{ and } \boldsymbol{\pi}.\mathbf{1} = 1. \tag{6.3}$$

For the PSAT model, denoting $K_b(m) = \{m : m - k_{b/m} = n\}$, at order n the conservation equation is:

$$\pi_n(\lambda + \sum_{b:N_b < n} f_b) = \pi_{n-1}\lambda + \sum_{b:N_b \leq n} f_b(\sum_{m \in K_b(n)} \pi_m), \tag{6.4}$$

Condition { $N_b < n$ } is required to go out of state n using line b. Condition { $N_b \leq n$ } stems from the requirement of $K_b(n) \neq \varnothing$ to come in state n from m such that $m - k_{b/m} = n$, which can hold only if $m \geq N_b + 1$ hence $n \geq N_b$.

Let us multiply in a formal way each side of eqn. (6.4) by ζ^n and sum over n. Then stochastic equilibrium amounts to the equality of two formal series:

$$\sum_n \pi_n (\lambda + \sum_{b:N_b<n} f_b)\zeta^n = \frac{\lambda}{\zeta}(\sum_n \pi_n \zeta^n) + \sum_b f_b \sum_{m>N_b} \pi_m \zeta^{m-k_{b/m}} . \quad (6.5)$$

This equation involves the generating function of distribution π, $N^*(\zeta) \equiv \sum_n \pi_n \zeta^n$. The transformation of the matrix conservation equation into the functional equation is useful if the functional equation admits a simple solution $N^*(\zeta)$. This holds notably when the transitions occur between neighbouring states and have homogeneous transition rates.

Instance 1 (continued). For the binary 1/1 model, any state $n > \nu$ is attained from $n-1$ with rate λ or $n+1$ with rate $f_Z \equiv f_a + f_b$ since both lines contribute to that transition; it is left towards $n+1$ at rate λ or towards $n-1$ at rate f_Z.The balance equation is:

$$\pi_n(\lambda + f_Z) = \pi_{n-1}\lambda + \pi_{n+1} f_Z . \quad (6.6a)$$

State $n = \nu$ exports to $\nu+1$ at rate λ and to $\nu-1$ at rate f_a; it imports from $\nu-1$ at rate λ or $\nu+1$ at rate f_Z, yielding:

$$\pi_\nu(\lambda + f_a) = \pi_{\nu-1}\lambda + \pi_{\nu+1} f_Z . \quad (6.6b)$$

State $n \in \{1,..\nu-1\}$ exports to $n+1$ at rate λ and to $n-1$ at rate f_a; it imports from $n-1$ at rate λ or $n+1$ at rate f_a, yielding:

$$\pi_n(\lambda + f_a) = \pi_{n-1}\lambda + \pi_{n+1} f_a . \quad (6.6c)$$

At $n = 0$, $$\pi_0 \lambda = \pi_1 f_a . \quad (6.6d)$$

It is shown in the Appendix that

$\forall n \le \nu$, $$\pi_n = \varphi^{\nu-n} \pi_\nu , \quad (6.7a)$$

$\forall m \ge 0$, $$\pi_{\nu+m} = \rho^m \pi_\nu , \quad (6.7b)$$

wherein $\rho \equiv \lambda / f_Z$ and $\varphi \equiv f_a / \lambda$. This yields that:

$$\pi_\nu = [\varphi \frac{1-\varphi^\nu}{1-\varphi} + \frac{1}{1-\rho}]^{-1}, \quad (6.8a)$$

$$N^*(\zeta) = \pi_\nu \{\varphi \frac{\varphi^\nu - \zeta^\nu}{\varphi - \zeta} + \frac{\zeta^\nu}{1-\rho\zeta}\} . \quad (6.8b)$$

Instance 2 (continued). In the binary $1/\infty$ model, any state $n > \nu$ exports to $n+1$ at rate λ and to $n-1$ at rate f_a and to ν at rate f_b; it imports from $n-1$ at rate λ or $n+1$ at rate f_a, yielding:

$$\pi_n(\lambda + f_Z) = \pi_{n-1}\lambda + \pi_{n+1} f_a . \quad (6.9a)$$

Any state $n \in \{1,..\nu\}$ behaves as a state $n \in \{1,..\nu-1\}$ in the 1/1 model, cf. (6.6c). State 0 still obeys to (6.6d). The formulae for the stationary probabilities, π_ν and $N^*(\zeta)$ are identical to those in the 1/1 model except for a specific definition of ρ (as the solution of $\rho f_a + \lambda / \rho = \rho + f_Z$, cf. Appendix).

In both binary models, (6.7) implies that $\nu = N_b$ is the mode of the stationary distribution.

6.3 Macroscopic properties

The generating function is endowed with some important properties, among which that

$$\mathrm{E}[X] = \sum_{n\geq 0} n\,\pi_n = \frac{\mathrm{d}}{\mathrm{d}\zeta}\mathrm{N}^*(\zeta)\Big|_{\zeta=1},$$

$$\mathrm{E}[X(X-1)] = \sum_{n\geq 0} n(n-1)\pi_n = \frac{\mathrm{d}^2}{\mathrm{d}\zeta^2}\mathrm{N}^*(\zeta)\Big|_{\zeta=1}.$$

Combining the two formulae yields the variance $\mathrm{V}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$.

By Little's law, the average number of customers in the station is equal to the average individual wait time multiplied by the flow rate of customers: $\overline{w}_Z = \mathrm{E}[X]/\lambda$.

Binary model. The formula for $\mathrm{E}[X]$ is established in the appendix:

$$\mathrm{E}[X] = \pi_\nu[\varphi\frac{\varphi^\nu - 1 + \nu(1-\varphi)}{(1-\varphi)^2} + \frac{\nu(1-\rho)+\rho}{(1-\rho)^2}]. \qquad (6.10a)$$

So $$\overline{w}_Z = \frac{\pi_\nu}{\lambda}[\varphi\frac{\varphi^\nu - 1 + \nu(1-\varphi)}{(1-\varphi)^2} + \frac{\nu(1-\rho)+\rho}{(1-\rho)^2}]. \qquad (6.10b)$$

When φ is small enough and ν is large enough, the average stock size tends to the attractivity threshold which is the mode of the stationary distribution:

$$\mathrm{E}[X] \approx \pi_\nu \nu[\frac{\varphi}{1-\varphi} + \frac{1}{1-\rho}] \approx \nu.$$

Conditional on state n, a vehicle of line z that arrives gets a passenger load of $k_{z/n}$. The line has passenger flow averaged over time instants (hence the stationary distribution):

$$\overline{x}_z = f_z \sum_n k_{z/n}\pi_n.$$

Instance 1 (continued). In the 1/1 binary model,

$$\overline{x}_a = f_a \sum_{n\geq 1}\pi_n = f_a(1-\pi_0) = f_a(1-\pi_\nu\varphi^\nu), \qquad (6.11a)$$

$$\overline{x}_b = f_b \sum_{n>\nu}\pi_n = f_b\,\pi_\nu\,\rho(1-\rho)^{-1}. \qquad (6.11b)$$

Instance 2 (continued). In the $1/\infty$ binary model, $\overline{x}_a$ is same as in (6.11a) whereas

$$\overline{x}_b = f_b \sum_{n>\nu}\pi_n k_{b/n} = f_b \pi_\nu \sum_{m\geq 1}\rho^m m = f_b\,\pi_\nu\,\rho(1-\rho)^{-2}. \qquad (6.12)$$

In both instances, letting $\varepsilon = 0$ under 1/1 or $\varepsilon = 1$ under $1/\infty$, the ratio of line flow between the two lines is:

$$\frac{\overline{x}_a}{\overline{x}_b} = \frac{f_a}{f_b}\frac{(1-\rho)^{1+\varepsilon}}{\rho}[\frac{1}{\pi_\nu} - \varphi^{-\nu}] = \frac{f_a}{f_b}(1-\rho)^\varepsilon[1 + \frac{1-\rho}{\rho}\frac{1-\varphi^\nu}{1-\varphi}]. \qquad (6.13)$$

As φ and ρ depend of f_a, f_b and λ the ratio of line flow cannot be equal to that of line frequency except on some special circumstance such as $\nu = 0$ under 1/1.

Lastly, the average run time and travel time are, respectively:

$$\bar{t}_Z = \frac{1}{\lambda}\sum_{b \in Z} \bar{x}_b t_b \,, \tag{6.14}$$

$$\bar{g}_Z = \bar{t}_Z + \alpha \bar{w}_Z \,. \tag{6.15}$$

In the binary instances,

$$\bar{t}_Z = \frac{1}{\lambda}[t_a f_a (1 - \pi_\nu \varphi^\nu) + t_b f_b \, \pi_\nu \, \rho (1-\rho)^{-1-\varepsilon}], \tag{6.16}$$

$$\bar{g}_Z = \frac{1}{\lambda}\left(\alpha\pi_\nu [\varphi \frac{\varphi^\nu - 1 + \nu(1-\varphi)}{(1-\varphi)^2} + \frac{\nu(1-\rho)+\rho}{(1-\rho)^2}] + t_a f_a (1-\pi_\nu \varphi^\nu) + t_b f_b \, \pi_\nu \, \rho(1-\rho)^{-1-\varepsilon} \right). \tag{6.17}$$

6.4 Numerical illustration

Let us come back to instance 1 as dealt with in Section 2. Recall that $f_a = f_b = 10$/h.

Fig. 3 depicts the variation of $\bar{x}_a$ with λ ranging from 0 to $f_Z = 20$/h, for two values $\nu = 3$ and $\nu = 5$ that correspond to Priority Queuing and Mingled Waiting, respectively. Fig. 4 (resp. 5) depicts the variation of line flow ratio $\bar{x}_a / \bar{x}_b$ (resp. $\bar{g}_Z$) with respect to λ in both cases.

It comes out that the average travel time increases more than linearly with demand volume λ: when λ reaches 19 i.e. 95% of the supplied capacity, the travel time is more than four times that in the absence of congestion. The faster line, a, accommodates most of the demand at the low volumes (Fig. 4). The flow share of the other line reaches ½ when demand approaches capacity, in accordance with the share of supplied capacity. MW is more conservative than PQ towards the faster line: this could be expected from Section 2.

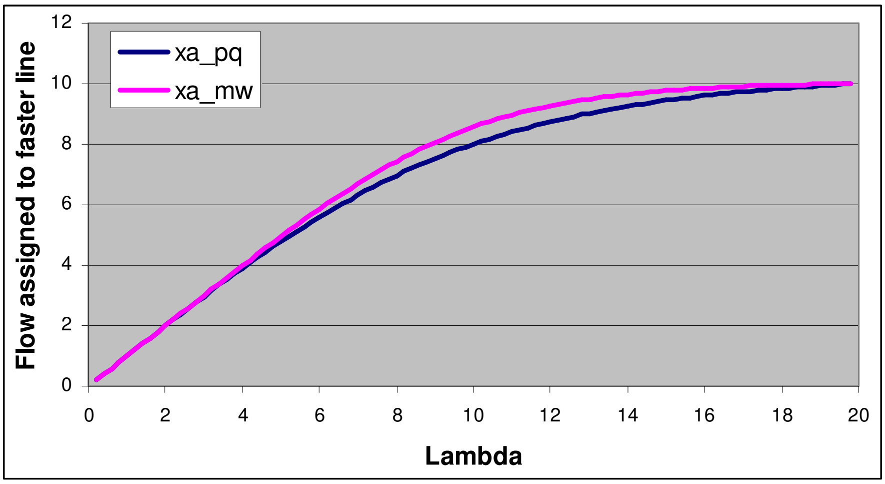

Fig. 3. Variation of $\bar{x}_a$ with λ under PQ ($\nu = 3$) and MW ($\nu = 5$).

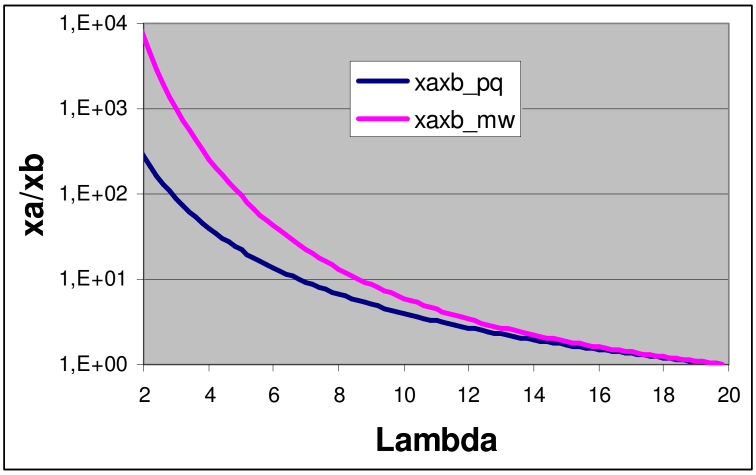

Fig. 4. Variation of $\bar{x}_a/\bar{x}_b$ ***with*** λ ***under PQ*** $(\nu = 3)$ ***and MW*** $(\nu = 5)$.

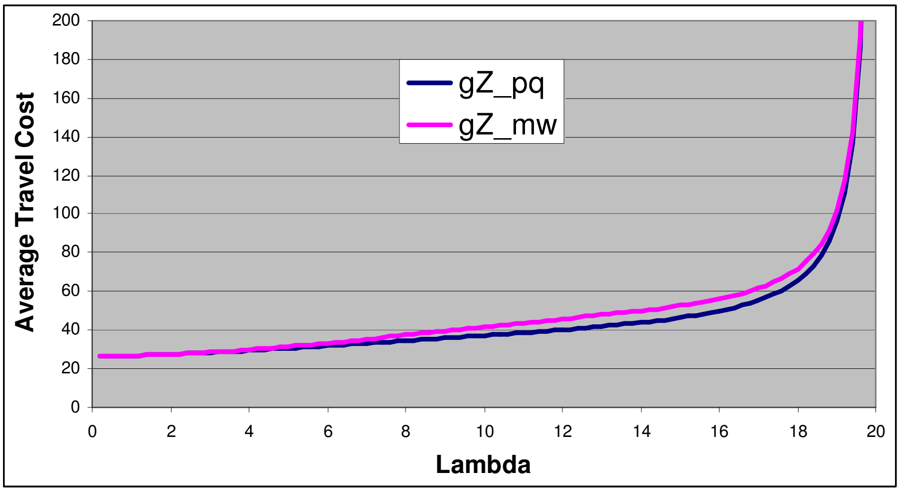

Fig. 5. Variation of $\bar{g}_Z$ ***with*** λ ***under PQ*** $(\nu = 3)$ ***and MW*** $(\nu = 5)$.

7. Conclusion

7.1 Summary

At a station platform, the waiting time for a passenger to board in a transit vehicle depends on not only the line frequency and capacity by vehicle but also the stock of passengers and their waiting discipline. These influence also route choice when there are several alternatives. A theory has been provided to represent passenger waiting and stock on platform together with line attractivity. Based on the characteristics of the line set, there is a maximal stock size up to which a given line is unattractive: this is the line attractivity threshold. When the stock exceeds the threshold, then the attractive capacity of a vehicle is the minimum of the supplied capacity and the rest of stock size minus threshold. Two waiting disciplines have been considered, priority queuing where better ranked passengers have better access versus mingled waiting; individual behaviour has been assumed under both disciplines – it turns out that priority queuing is more advantageous to the passengers as it yields system optimization beyond user equilibrium.

Composition rules have been provided to represent the evaluation of route options conditional on stock size by the individual passenger, yielding the average travel time of a line bundle. Conditions for line attractivity and for user equilibrium have been stated; the existence and uniqueness of an attractive line bundle which yields user equilibrium have been demonstrated, on the basis of a recursive structure of the attractive set at a given stock size.

The resulting model is basically stationary: it involves three traffic variables namely stock size, travel time of line bundle and exit flow. So it amounts to a fundamental traffic diagram for a line bundle: the consequences for transit assignment have been explored. Lastly, a Markovian model has been developed to characterize the stationary state of the system under a given flow of customer arrivals: analytical solutions have been provided for binary models where lines have unit or infinite vehicle capacity.

7.2 Research perspectives

The scope of the model is limited to one destination, homogeneous passengers and a station platform organized into one boarding zone only (Leurent, 2009a). Specific work has been invested by the author and co-workers to extend the model to continuous variables (Leurent, 2010b), more sophisticated station layout (Chandakas and Leurent, 2010), several destinations (Leurent et al, 2011).

Further research could be aimed at a dynamic version in which the passenger stock would vary over time; at the development of robust behavioural rules about the evaluation of line characteristics by the individual passenger; at the inclusion of stochastic features such as variations in vehicle capacity among the runs that service a route, or clustered arrivals of passengers – e.g. at a transfer station.

8. References

Cepeda M., R. Cominetti and M. Florian, A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria, *Transportation Research Part B*, Vol. 40, 2006, pp. 437–459.

Chandakas E and Leurent F (2010). *Le modèle PSAT à une ou deux zones d'embarquement*. Working Document, December 2010.

Chriqui C. and P. Robillard, Common bus lines, Transportation Science, Vol. 9, 1975, pp. 115–121.

Cominetti R, Correa J (2001) Common-lines and passenger assignment in congested transit networks. *Transportation Science* **35**: 250-267.

De Cea J, Fernandez E (1993) Transit assignment for congested public transport systems: an equilibrium model. *Transportation Science* **27**: 133-147.

Greenshields BD (1935) A study of traffic capacity. *Publ. Rds*, Washington, **23**, pp. 57 sq.

Hamdouch Y., P. Marcotte and S. Nguyen (2004) Capacitated transit assignment with loading priorities, *Mathematical Programming B*, Vol. 101, pp. 205–230.

HCM (2010). *Highway Capacity Manual*. See article in http://en.wikipedia.org/wiki/Highway_Capacity_Manual

Kleinrock L (1975) *Queueing systems*. Volume I: Theory. Wiley, New York.

Kurauchi F., M. G. H. Bell and J.D. Schmöcker (2003) Capacity constrained transit assignment with common lines, *Journal of Mathematical Modelling and Algorithms*, 2, 309-327.

Leurent F (2009a) *Contraintes de capacité et attractivité des services de transport collectif : modélisation stochastique*. Working Document, December 17, 2009.

Leurent F (2009b) *Choix d'itinéraire en transport collectif sous contrainte de capacité : modèles d'optimisation continue*. Working Document, December 24, 2009.

Leurent F and Askoura Y (2010). *The person capacity of a transit route: a review, assessment and benchmark of static models for network traffic assignment*. Paper presented at the TRB Annual Congress 2010, reference TRB # 10-2020.

Leurent F, Chandakas E, Poulhès A (2011) *User and service equilibrium in a structural model of traffic assignment to a transit network*. Paper proposed at the 2011 Conference of the EWGT, Poznan, September.

Shimamoto H., Y. Iida, J. D. Schmöcker, F. Kurauchi and M. G. H. Bell (2005) Evaluating public transit congestion mitigation measures using a passenger assignment model, *Journal of the Eastern Asia Society for Transportation Studies*, 6, 2076-2091.

Spiess H, Florian M (1989) Optimal strategies: a new assignment model for transit networks. *Transportation Research B* 23: 83-102.

9. Appendix

9.1 PQ as System optimization for MW

Let us show that the marginal overall travel time under MW mingled waiting is identical to the individual travel time under PQ for the last passenger in the stock by finishing the demonstration in Section 2.5.

If a line a not in $A(n-1)$ is included in $A(n)$ by comparison to a criterion (2.34) that is based on previous determinations, the criterion is the same for both PQ and MW social. The added line turns $n\bar{g}^{(n)}_{\mathrm{MW}}$ into

$$n\tilde{g}^{(n)}_{\mathrm{MW}} = \frac{f_a}{f_{A(n)+a}}(t_a + (n-1)\bar{g}^{(n-1)}_{\mathrm{MW}}) + \frac{f_{A(n)}}{f_{A(n)+a}} n\bar{g}^{(n)}_{\mathrm{MW}}$$

By convex combination, the $(n-1)\bar{g}^{(n-1)}_{\mathrm{MW}}$ part within $n\bar{g}^{(n)}_{\mathrm{MW}}$ is maintained. Thus

$$n\tilde{g}^{(n)}_{\mathrm{MW}} - (n-1)\bar{g}^{(n-1)}_{\mathrm{MW}} = \frac{f_a}{f_{A(n)+a}} t_a + \frac{f_{A(n)}}{f_{A(n)+a}} \bar{g}^{(n)}_{\mathrm{PQ}}.$$

On the other side, under PQ, $\tilde{g}^{(n)}_{\mathrm{PQ}} = \frac{f_a}{f_{A(n)+a}} t_a + \frac{f_{A(n)}}{f_{A(n)+a}} \bar{g}^{(n)}_{\mathrm{PQ}}$.

Thus $n\tilde{g}^{(n)}_{\mathrm{MW}} - (n-1)\bar{g}^{(n-1)}_{\mathrm{MW}} = \tilde{g}^{(n)}_{\mathrm{PQ}}$ and the equivalency between the two routing behaviours is maintained by the inclusion of another line as well as by order incrementation.

9.2 Inductive solution of binary Markovian model

Let us search for a geometric solution:

$$\pi_{\nu+m} = \rho^m \pi_\nu \,,\; m \geq 0 \,. \tag{9.1}$$

Under 1/1, (6.6a) at $\nu + m + 1$ requires that $\rho(\lambda + f_Z) = \lambda + \rho^2 f_Z$ i.e. $(\rho^2 - \rho) f_Z = (\rho - 1)\lambda$ or $\rho = \lambda / f_Z$ which must be less than 1 to preserve stationarity (demand λ less than supplied capacity f_Z).

Under $1/\infty$, (6.9a) at $\nu + m + 1$ requires that $\rho(\lambda + f_Z) = \lambda + \rho^2 f_a$. The discriminant of this second order equation is $\Delta = (\lambda + f_Z)^2 - 4\lambda f_a = (\lambda - f_Z)^2 + 4\lambda f_b$ which is ≥ 0. Solution has the form $\rho^* = \frac{1}{2 f_a}[\lambda + f_Z + \eta\sqrt{\Delta}]$ with $\eta \in \{-1, +1\}$: only the positive value $+1$ is likely for η.

Whatever the waiting discipline, for $0 < n < \nu$, (6.6c) is equivalent to

$$(\pi_{n-1} - \pi_n)\lambda = f_a (\pi_n - \pi_{n+1}) \,. \tag{9.2}$$

Let $\delta_n \equiv \pi_{n-1} - \pi_n$ and $\varphi \equiv f_a / \lambda$. Then (9.2) implies that

$$\delta_n = \varphi^{\nu - n} \delta_\nu \,, \tag{9.3a}$$

$$\pi_n = \pi_\nu + \delta_\nu \frac{1 - \varphi^{\nu - n}}{1 - \varphi} \,, \tag{9.3b}$$

since $\pi_n = (\sum_{i=n}^{\nu-1} \pi_i - \pi_{i+1}) + \pi_\nu = \pi_\nu + \sum_{i=n+1}^{\nu} \delta_i = \pi_\nu + \delta_\nu \sum_{i=n+1}^{\nu} \varphi^{\nu - i} = \pi_\nu + \delta_\nu \sum_{i=0}^{\nu-n-1} \varphi^i$.

Under 1/1, (6.6b) at $n = \nu$ yields that $\pi_{\nu-1} = \varphi \pi_\nu$ hence $\delta_\nu = \pi_\nu(\varphi - 1)$ so

$$\pi_n = \pi_\nu \varphi^{\nu - n} \text{ for } n \leq \nu \,. \tag{9.4}$$

Similarly, under $1/\infty$ eqn. (6.6c) applied to $n = \nu$ implies that $\pi_\nu(\lambda + f_a) = \pi_{\nu-1}\lambda + \pi_\nu \rho f_Z$ hence $\pi_{\nu-1} = \varphi \pi_\nu$ yielding also (9.4).

Lastly, the value of π_ν stems from the law of total probability:

$$1 = \sum_{n \geq 0} \pi_n = \pi_\nu [\sum_{n=0}^{\nu} \varphi^{\nu - n} + \sum_{m \geq 1} \rho^m] = \pi_\nu [\varphi^\nu \frac{1 - \varphi^{-\nu-1}}{1 - \varphi^{-1}} + \frac{\rho}{1 - \rho}] \text{, yielding}$$

$$\pi_\nu = [\frac{1 - \varphi^{\nu+1}}{1 - \varphi} + \frac{\rho}{1 - \rho}] \,. \tag{9.5}$$

9.3 Generating function

$$\begin{aligned} N^*(\zeta) &\equiv \sum_{n \geq 0} \pi_n \zeta^n \\ &= \pi_\nu \{\varphi^\nu \sum_{n=0}^{\nu-1} (\frac{\zeta}{\varphi})^n + \zeta^\nu \sum_{i=0}^{\infty} \rho^i \zeta^i\} \\ &= \pi_\nu \{\varphi^\nu \frac{1 - (\zeta/\varphi)^\nu}{1 - \zeta/\varphi} + \frac{\zeta^\nu}{1 - \rho\zeta}\} \end{aligned}$$

Thus

$$N^*(\zeta) = \pi_\nu \{\varphi \frac{\varphi^\nu - \zeta^\nu}{\varphi - \zeta} + \frac{\zeta^\nu}{1-\rho\zeta}\} \tag{9.6}$$

From this stems:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\zeta} N^*(\zeta) &= \pi_\nu [\varphi \frac{-\nu\zeta^{\nu-1}(\varphi-\zeta)+\varphi^\nu-\zeta^\nu}{(\varphi-\zeta)^2} + \frac{\nu\zeta^{\nu-1}(1-\rho\zeta)+\rho\zeta^\nu}{(1-\rho\zeta)^2}] \\ &= \pi_\nu [\varphi \frac{\varphi^\nu+(\nu-1)\zeta^\nu-\varphi\nu\zeta^{\nu-1}}{(\varphi-\zeta)^2} + \frac{\nu\zeta^{\nu-1}-(\nu-1)\rho\zeta^\nu}{(1-\rho\zeta)^2}]\end{aligned} \tag{9.7}$$

By letting ζ tend to 1, we obtain that

$$\mathrm{E}[X] = \pi_\nu [\varphi \frac{\varphi^\nu - 1 + \nu(1-\varphi)}{(1-\varphi)^2} + \frac{\nu(1-\rho)+\rho}{(1-\rho)^2}]. \tag{9.8}$$

When φ is small enough and ν is large enough, approximately

$$\mathrm{E}[X] \approx \pi_\nu \nu [\frac{\varphi}{1-\varphi} + \frac{1}{1-\rho}] \approx \nu .$$

Thus the average stock size tends to the mode of the stationary distribution.

9.4 Line flow of second line

Under 1/1, $k_{b/n} = 1$ if $n > \nu$ or zero otherwise so

$$\begin{aligned}\bar{x}_b &= f_b \textstyle\sum_{n>\nu} \pi_n k_{b/n} = f_b \pi_\nu \sum_{m\geq 1} \rho^m \\ &= f_b \pi_\nu \frac{\rho}{1-\rho}\end{aligned}$$

Under $1/\infty$, $k_{b/n} = (n-\nu)^+$ so

$$\begin{aligned}\bar{x}_b &= f_b \textstyle\sum_{n>\nu} \pi_n k_{b/n} = f_b \pi_\nu \sum_{m\geq 1} \rho^m m \\ &= f_b \pi_\nu \rho \textstyle\sum_{m\geq 0} m \rho^{m-1} \\ &= f_b \pi_\nu \frac{\rho}{(1-\rho)^2}\end{aligned}$$